

**Estimation de la fonction de régression avec des pertes
asymétriques**

par

Marwa Hamdi

Mémoire présenté au Département de mathématiques
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, juin 2020

À mes chers parents ainsi qu'à mes frères et soeurs.

Le 23 juin 2020,

*le jury a accepté le mémoire de Madame Marwa Hamdi
dans sa version finale.*

Membres du jury

Professeur Taoufik Bouezmarni

Directeur de recherche

Département de mathématiques

Professeur Patrick Richard

Membre externe

Département d'économie

Professeur Bernard Colin

Président-rapporteur

Département de mathématiques

RÉSUMÉ

La régression linéaire basé sur la perte quadratique est un outil statistique qui permet de modéliser l'impact des variables explicatives sur la moyenne conditionnelle de la variable d'intérêt. L'estimation de ses paramètres par la méthode des moindres carrés ordinaire (MCO) est très populaire. Cette méthode produit des estimateurs efficaces lorsque l'erreur est normale, mais elle souffre de plusieurs problèmes, notamment l'inefficacité de l'estimation au-delà de la moyenne et la non-robustesse des estimateurs. Plusieurs méthodes ont été proposées pour résoudre ces problèmes telles que la régression quantile (RQ) et la régression expectile (RE).

Dans ce mémoire, nous introduisons une nouvelle approche robuste et efficace que nous allons appeler la régression Hamdi et Bouezmarni (HB). Elle permet d'étudier l'influence des régresseurs sur la distribution de la variable réponse dans le cas des données complètes et censurées. Au niveau de l'efficacité et de la robustesse, la performance de la nouvelle méthode d'estimation de paramètres, se situe entre celle de la méthode robuste (RQ) et la méthode efficace (RE). Dans ce travail, les définitions sont présentées et les propriétés asymptotiques de la fonction de régression avec différentes pertes symétriques et asymétriques sont étudiées. Par simulations, les méthodes RQ, RE et HB sont comparées.

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude et mes sincères remerciements à mon directeur de maîtrise le Professeur T. Bouezmarni. Il a fait preuve de dévouement et s'est toujours montré très disponible, ce que j'ai fortement apprécié. Son aide et ses précieux conseils m'ont été indispensables durant ces dernières années, ce qui m'a permis de réaliser ce travail dans de bonnes conditions.

Je remercie les membres de mon jury, Monsieur B. Colin et Monsieur P. Richard pour avoir accepté d'évaluer ce travail. Je remercie également tous mes professeurs, mes collègues et tout l'ensemble du personnel du Département de mathématiques de l'Université de Sherbrooke.

Mes remerciements vont aussi à ma famille, mes amis et toutes les personnes qui m'ont accompagnée et soutenue tout au long de ces années d'études. Je pense tout particulièrement à mes chers parents ainsi qu'à mes frères et soeurs.

Marwa Hamdi
Sherbrooke, juin 2020

TABLE DES MATIÈRES

RÉSUMÉ	iv
REMERCIEMENTS	v
TABLE DES MATIÈRES	vi
LISTE DES FIGURES	ix
LISTE DES TABLES	xi
INTRODUCTION	1
CHAPITRE 1 — Régression pour les données complètes	4
1.0.1 Modélisation	5
1.1 Régression linéaire basée sur les pertes symétriques	6
1.1.1 Régression linéaire moyenne	6
1.2 La méthode Lasso	10

1.2.1	Interprétation géométrique	11
1.2.2	Régression linéaire robuste	11
1.2.3	Régression linéaire médiane (LAD)	19
1.3	Régression linéaire basée sur les pertes asymétriques	22
1.4	Fonctions de perte asymétriques	22
1.4.1	Quantiles	22
1.4.2	Expectiles	25
1.4.3	Relation entre les quantiles et les expectiles	27
1.5	Régression quantile et régression expectile	28
1.5.1	Régression quantile linéaire	28
1.5.2	Régression expectile linéaire	35
1.6	Méthode lasso pour la régression expectile	42
1.7	Condition d’optimalité	43
CHAPITRE 2 — Régression linéaire pour les données censurées		44
2.1	Distribution de la durée de survie	45
2.1.1	Censure à droite	46
2.1.2	Estimateur de Kaplan-Meier	48
2.1.3	Estimateur de Kaplan-Meier généralisé	52
2.2	Estimation de la régression quantile et expectile : cas des données censurées	55
2.2.1	Estimation de la régression quantile	57

2.2.2	Estimation de la régression expectile	59
CHAPITRE 3 — Approche robuste et efficace pour l'estimation de la		
	régression linéaire	64
3.1	Introduction	64
3.2	Estimation HB de la régression linéaire pour les données complètes . . .	65
3.2.1	Estimateur HB	66
3.2.2	Propriétés asymptotiques	68
3.3	Estimation HB de la régression linéaire pour les données censurées	71
3.3.1	Estimation de la régression HB	72
CHAPITRE 4 — Simulations		74
4.1	Les Modèles et leurs paramètres	75
4.2	Les résultats	78
4.2.1	Normalité de l'estimateur HB	78
4.2.2	Comparaison des estimateurs	81
4.2.3	Discussion	88
CONCLUSION		89
BIBLIOGRAPHIE		91

LISTE DES FIGURES

1.1	Cas où la solution se trouve dans la zone qui annule le coefficient β_2	11
1.2	Différents points aberrants.	13
1.3	Estimateur par MCO avec des données aberrantes.	14
1.4	Fonctions objectives pour MC, Huber ($k = 1,345$) et Tukey ($k = 4,685$).	16
1.5	Fonctions de perte associées à la moyenne et la médiane.	19
1.6	La régression médiane avec un point de levier.	21
1.7	Fonction de perte des quantiles.	24
1.8	Fonction de perte des expectiles.	26
1.9	Expectiles et quantiles de la distribution normale.	28
1.10	La régression quantile pour différents τ	30
1.11	Illustration de la régression expectile pour différents ω avec le modèle $Y = \beta_0^\omega + \beta_1^\omega x_{i,1} + \alpha x_{i,1} \epsilon_i$	36
2.1	Exemple de censure à droite aléatoire.	48

2.2	Une illustration des estimateurs empiriques et Kaplan-Meier de la distribution.	52
2.3	Les noyaux d'Epanechnikov et "Triweight",	54
2.4	Illustration d'une relation linéaire entre la variable réponse censurée et une variable explicative. Trois méthodes d'estimation : MCO et la régression quantile avec la correction (les droites continues) et la méthode MCO sans correction (les droites discontinues).	62
3.1	Fonctions de perte $\zeta_{1,5}^k$ avec différentes valeurs de k	67
3.2	Relation linéaire entre mpg et wt avec trois méthodes d'estimation : MCO, régression quantile et régression HB. Sans valeurs aberrantes, les droites en lignes continues et avec points aberrants les droites sont discontinues. Les deux points oranges sont des points aberrants.	68
4.1	Densité de l'estimateur HB du modèle 1 pour de différentes valeurs de τ où $n = 200$	79
4.2	Densité de HB du modèle 3 pour différentes valeurs de τ où $n = 200$. . .	79
4.3	Densité de l'estimateur HB du modèle 5 pour de différentes valeurs de τ où $n = 200$ dans le cas de censure à droite où $p = 0, 25$	80
4.4	Densité de l'estimateur HB du modèle 5 pour de différentes valeurs de τ où $n = 200$ dans le cas de censure à droite où $p = 0, 5$	80

LISTE DES TABLEAUX

1.1	Exemples des fonctions objectifs utilisées pour les M-estimateurs, $k \geq 0$	15
4.1	Table présentant le carré du biais (10^{-3}) et les écarts types (10^{-2}) obtenus à partir du modèle 1 de chacune des trois méthodes pour $n = 50, 100$ et 200 et $\tau = 0, 25, 0, 5$ et $0, 75$.	82
4.2	Table présentant le carré du biais (10^{-3}) et les écarts types (10^{-2}) obtenus à partir du modèle 2 de chacune des trois méthodes pour $n = 50, 100$ et 200 et $\tau = 0, 25, 0, 5$ et $0, 75$.	83
4.3	Table présentant le carré du biais (10^{-3}) et les écarts types (10^{-2}) obtenus à partir du modèle 3 de chacune des trois méthodes pour $n = 50, 100$ et 200 et $\tau = 0, 25, 0, 5$ et $0, 75$.	84
4.4	Table présentant le carré du biais et les écarts types (10^{-2}) obtenus à partir du modèle 4 de chacune des trois méthodes pour $n = 50, 100$ et 200 et $\tau = 0, 25, 0, 5$ et $0, 75$.	85
4.5	Table présentant le carré du biais ($\div 10^5$) et les écarts types (10^{-2}) obtenus à partir du modèle 5 de chacune des trois méthodes pour $n = 50, 100$ et 200 , $\tau = 0, 25, 0, 5$ et $0, 75$ et $p = 0, 25$.	86

4.6	Table présentant le carré du biais ($\div 10^5$) et les écarts types (10^{-2}) obtenus à partir du modèle 5 de chacune des trois méthodes pour $n = 50, 100$ et 200 , $\tau = 0, 25, 0, 5$ et $0, 75$ et $p = 0, 5$	87
-----	---	----

INTRODUCTION

L'efficacité et la robustesse des estimateurs sont les deux concepts les plus recherchés en statistique mathématique, entre autres, dans le cas de la régression linéaire. Cette dernière a pour but d'étudier la relation entre la variable réponse et les variables explicatives en supposant un lien linéaire, entre la variable réponse et les variables explicatives, contrôlé par des paramètres inconnus. L'estimation de ces paramètres est basée principalement sur une fonction de perte. La méthode des moindres carrés ordinaire (MCO) est l'une des premières méthodes développée pour estimer les paramètres de la régression linéaire. En effet, elle estime l'effet des régresseurs sur la moyenne de la distribution de la variable d'intérêt, en se basant sur une perte quadratique (la somme des carrés des résidus). L'estimation par cette méthode est efficace, mais elle est très sensible aux points aberrants. En d'autres termes, l'estimateur de cette méthode est non robuste. Afin de résoudre les limitations de la MCO, plusieurs méthodes de régression avec différentes fonctions de perte ont été développées, notamment la régression linéaire robuste. Cette dernière considère des fonctions de perte qui sont quadratiques pour les petites valeurs des erreurs et une perte linéaire ou constante pour les grandes valeurs des erreurs. Ceci permet de détecter les points aberrants et leur affecter un poids important. En utilisant des pertes adaptées, des estimateurs robustes ont été proposés et étudiés. La régression médiane est une alternative robuste à la régression moyenne du fait de l'utilisation de la perte absolue au lieu

de la perte quadratique de la méthode MCO. Les régressions linéaires moyenne, médiane et la régression linéaire robuste utilisent des poids symétriques et modélisent la tendance centrale de la variable d'intérêt conditionnelle aux variables explicatives. Cependant, ces régressions ne reflètent pas toutes les informations sur le comportement de la variable réponse conditionnelle aux régresseurs (variables explicatives). Par exemple, on ne peut pas modéliser les petites valeurs de la variable réponse. Pour résoudre ce problème, des méthodes de régression avec des fonctions de perte asymétriques capables de modéliser complètement la distribution conditionnelle ont vu le jour. Parmi ces méthodes, on trouve la régression quantile (RQ) et la régression expectile (RE). Ces deux régressions sont des généralisations des régressions moyennes et médianes. La RQ modélise les quantiles tandis que la RE modélise les expectiles de la distribution de la variable d'intérêt en fonction des variables explicatives. Les deux modèles prennent en considération le comportement de la variable d'intérêt en fonction des variables explicatives du modèle. Dans le présent mémoire, nous présentons toutes ces méthodes avec leurs propriétés asymptotiques et ceci dans deux cas : le cas d'une variable réponse complètement observée et l'autre cas d'une variable d'intérêt soumise à de la censure à droite.

Les RQ et RE jouent des rôles très similaires dans la modélisation, mais modélisent deux concepts différents. Elles se distinguent principalement par rapport à la robustesse et l'efficacité des estimateurs. La RQ est très robuste, mais moins efficace par rapport au MCO et inversement, la RE qui est très efficace, mais moins robuste. La RQ hérite des propriétés de la régression médiane tandis que la RE hérite de celles de la régression moyenne. Dans ce travail, nous introduisons une nouvelle méthode de régression linéaire qui est à la fois efficace et robuste. Précisément, cette méthode est, plus efficace que la RQ, mais moins que la RE, et plus robuste que la RE, mais plus sensible que la RQ. Nous appelons ce modèle la régression de Hamdi et Bouezmarni(HB). L'étude des propriétés asymptotiques de cette nouvelle approche et sa performance, comparé à la RE et RQ,

sont présentées dans le cas des données complètes et censurées.

Le premier chapitre de ce mémoire présente les différentes méthodes d'estimation de la régression linéaire basées sur différentes fonctions de perte, symétriques et asymétriques, dans le cas des données complètes. Pour chaque méthode utilisée dans ce chapitre, nous présentons la définition, les hypothèses et les propriétés asymptotiques des estimateurs des paramètres du modèle linéaire. Le deuxième chapitre est consacré à la modélisation de la régression linéaire dans le cas des données censurées ; plus précisément, on suppose que la variable d'intérêt est censurée à droite, mais que les variables explicatives sont complètement observées. Nous rappelons des notions préliminaires, mais très importantes dans l'analyse de survie telles que les fonctions utiles pour modéliser le temps de survie, la censure à droite et l'estimateur de Kaplan-Meier. Dans la deuxième partie de ce chapitre, nous introduisons l'estimation de la régression avec différentes fonctions de perte, notamment la RQ et RE. Le chapitre trois est dédié à la nouvelle méthode de modélisation de la régression linéaire : la régression HB. Dans la première partie, nous présentons l'estimation et les propriétés asymptotiques de l'estimateur de la régression linéaire HB dans le cas des données complètes. Dans la deuxième partie, nous introduisons dans le cas des données censurées, l'estimateur de la régression HB et quelques résultats théoriques. Finalement, le chapitre quatre de ce mémoire est consacré à la comparaison des trois méthodes d'estimation, RQ, RE et HB par le biais de simulations. Nous concluons à la fin avec une discussion générale et quelques idées à développer dans le futur pour la régression HB.

CHAPITRE 1

Régression pour les données complètes

Les modèles de régression étudient la relation entre la variable dite d'intérêt Y et une ou plusieurs variables explicatives. Ces modèles permettent ensuite d'expliquer un phénomène déjà observé ou donner une prédiction de Y avec de nouvelles observations des variables explicatives. Parmi ces modèles on trouve la régression linéaire qui explique l'impact des variables explicatives $\mathbb{X} = (X_1, \dots, X_p) \in \mathbb{R}^p$ sur la variable d'intérêt $Y \in \mathbb{R}$ en supposant une relation linéaire. L'estimation de ce modèle se base principalement sur une fonction de perte. Dans ce chapitre, nous allons étudier plusieurs méthodes d'estimation de la régression linéaire, en se basant sur les différentes fonctions de perte (symétriques et asymétriques). Pour chaque méthode utilisée, nous présentons les estimateurs et leur perte associée ainsi que quelques propriétés asymptotiques des estimateurs.

1.0.1 Modélisation

Il existe plusieurs modèles de régression : paramétrique, semi-paramétrique et non-paramétrique. Parmi les modèles paramétriques, on trouve la régression linéaire qui permet de modéliser le lien entre un vecteur aléatoire de prédicteurs $\mathbb{X} = (X_1, \dots, X_p) \in \mathbb{R}^p$ et la variable réponse $Y \in \mathbb{R}$, en supposant une relation linéaire entre ces deux derniers. Ce modèle est de la forme suivante :

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon,$$

où $(\beta_0, \beta_1, \dots, \beta_p)$ est le vecteur des paramètres du modèle. Ce dernier est inconnu et le but de ce travail est d'estimer les valeurs des $(p + 1)$ paramètres du vecteur $(\beta_0, \beta_1, \dots, \beta_p)$ à partir d'un échantillon indépendant et identiquement distribué (i.i.d.) $\{(y_i, x_{i,1}, \dots, x_{i,p}), i = 1, \dots, n\}$ généré à partir du modèle linéaire.

$$y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p} + \epsilon_i, \quad i = 1, \dots, n \quad (1.1)$$

où :

- y_i est la i – ème observation de la variable Y ;
- $x_{i,j}$ est la i – ème observation de la j – ème variable explicative X_j ;
- ϵ_i est l'erreur du modèle ; il résume les informations manquantes qui permettraient d'expliquer linéairement les valeurs de y_i à l'aide des $x_{i,j}$, $j = 1, \dots, p$.

Pour estimer les valeurs des $(p + 1)$ paramètres du vecteur $(\beta_0, \beta_1, \dots, \beta_p)$ à partir d'un échantillon de n observations, on écrit le modèle linéaire, sous la forme matricielle pour simplifier les notations.

Définition 1.1 *Le modèle de la régression linéaire (1.1) s'exprime sous la forme suivante :*

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1.2)$$

où

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{x}_1^t \\ \mathbf{x}_2^t \\ \vdots \\ \mathbf{x}_n^t \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \text{ et } \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

1.1 Régression linéaire basée sur les pertes symétriques

L'estimation des paramètres du modèle (1.2) est toujours basée sur la minimisation des erreurs. Pour cela, on utilise une fonction de perte qui mesure l'erreur "globale" du modèle. Il existe plusieurs fonctions de perte et nous allons présenter les estimateurs basés sur les pertes symétriques.

1.1.1 Régression linéaire moyenne

Nous commençons par la méthode d'estimation des paramètres la plus populaire. L'estimation par les moindres carrés ordinaires est basée sur la perte quadratique. On considère le modèle (1.2), avec les hypothèses suivantes :

Hypothèses

H.1 : $\mathbb{R}g(\mathbf{X}) = p + 1$ (la matrice \mathbf{X} est pleine de rang).

H.2 : $\mathbb{E}(\epsilon_i) = 0, i = 1, \dots, n$.

H.3 : $\text{Var}(\epsilon_i) = \sigma^2, i = 1, \dots, n$ (homoscedasticité).

H.4 : les erreurs, ϵ_i , sont non corrélées.

H.5 : les ϵ_i sont non corrélées avec les variables explicatives, $\forall i = 1, \dots, n$.

Dans la régression moyenne, on cherche à modéliser l'espérance de Y conditionnelle à

X_1, \dots, X_p , c-à-d $\mathbb{E}(Y|X_1, \dots, X_p)$. En utilisant l'hypothèse H_2 , cette moyenne conditionnelle est donnée par :

$$\mathbb{E}(Y|X_1 = x_1, \dots, X_p = x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

La méthode des moindres carrés ordinaires demeure une des approches les plus utilisées qui permet d'obtenir des estimateurs de β non biaisés. Cette méthode est basée sur la minimisation de l'erreur quadratique moyenne.

Définition 1.2 (*Estimateur MCO*) Soit $\mathbf{X} \in M_{n \times (p+1)}$ de rang $p+1$. De plus, soit $\mathbf{Y} \in R^n$. L'estimation par MCO, basée sur un échantillon i.i.d. $\{(y_i, x_{i,1}, \dots, x_{i,p}), i = 1, \dots, n\}$, noté $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ des coefficients de régression $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ est donnée par,

$$\hat{\beta}^{MCO} = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p}))^2.$$

L'estimateur des MCO peut s'écrire sous la forme matricielle :

$$\hat{\beta}^{MCO} = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{X}\beta\|_2^2.$$

La condition H_1 , qui suppose que \mathbf{X} est de rang $p+1$, garantie que $\mathbf{X}^t \mathbf{X}$ est inversible, où \mathbf{X}^t est la matrice transposée de \mathbf{X} , et donc $\hat{\beta}^{MCO}$ est unique. En utilisant le fait que les dérivées partielles de $\|\mathbf{Y} - \mathbf{X}\beta\|_2^2$ par rapport aux composantes de β sont égales à zéro, on peut trouver la forme explicite de l'estimateur par MCO

$$\hat{\beta}^{MCO} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

Remarque 1.3 L'estimateur des MCO est égal à l'estimateur par la méthode du maximum de vraisemblance si on suppose que la variable réponse suit une loi normale de

moyenne μ et de variance σ^2 . En posant $\mu_i = \mathbf{x}_i^t \boldsymbol{\beta}$, la vraisemblance est donnée par

$$\mathbb{L}(\boldsymbol{\beta}, \sigma | \mathbf{y}) \propto \sigma^{-n} \prod_{i=1}^n \exp \left(-\frac{(y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2}{\sigma^2} \right).$$

Dans la section suivante, on présente quelques propriétés de l'estimateur par la méthode des MCO.

Propriétés asymptotiques

Proposition 1.4 *Sous les hypothèses H.1-H.5,*

1. $\hat{\boldsymbol{\beta}}^{MCO}$ est un estimateur sans biais, autrement dit $\mathbb{E}(\hat{\boldsymbol{\beta}}^{MCO}) = \boldsymbol{\beta}$.
2. La matrice de variance-covariance de $\hat{\boldsymbol{\beta}}^{MCO}$ est égale à $\text{Var}(\hat{\boldsymbol{\beta}}^{MCO}) = \sigma^2 \mathbb{E}(\mathbf{X}^t \mathbf{X})^{-1}$.
3. L'estimateur $\hat{\boldsymbol{\beta}}^{MCO}$ des MCO est de variance minimale parmi les estimateurs linéaires sans biais, noté BLUE (Best Linear Unbiased Estimator) de $\boldsymbol{\beta}$.

Maintenant, pour établir la convergence de l'estimateur des MCO ainsi que sa normalité, notons que l'estimateur peut s'écrire sous la forme :

$$\hat{\boldsymbol{\beta}}^{MCO} = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i y_i \right).$$

Proposition 1.5 *On considère le modèle (1.2) et $\hat{\boldsymbol{\beta}}^{MCO}$ l'estimateur par la méthode MCO. Alors, $\hat{\boldsymbol{\beta}}^{MCO}$ converge en probabilité vers $\boldsymbol{\beta}$.*

Preuve : *En effet en vertu de la loi faible des grands nombres on a*

$$\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \xrightarrow{P} \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t), \quad \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i y_i \xrightarrow{P} \mathbb{E}(\mathbf{x}_i y_i).$$

Or, d'après (1.2), $\mathbf{x}_i y_i = \mathbf{x}_i \mathbf{x}_i^t \boldsymbol{\beta} + \mathbf{x}_i \epsilon_i$. Donc, en appliquant l'espérance et en supposant que $\mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t)$ est non singulière, on en déduit la convergence de $\hat{\boldsymbol{\beta}}^{MCO}$ vers $\boldsymbol{\beta}$.

■

Le théorème suivant établit la normalité asymptotique de l'estimateur par les MCO. Ce résultat permet de construire les intervalles de confiance et faire des tests d'hypothèse sur les paramètres du modèle.

Théorème 1.6 *Sous les hypothèses H.1-H.5, on a :*

$$\sqrt{n}\Sigma^{1/2}(\hat{\beta}^{MCO} - \beta) \xrightarrow{d} N(0, \sigma^2 \mathbb{I}_{(p+1)(p+1)})$$

$$\text{où } \Sigma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t.$$

Preuve En effet, en utilisant la loi faible des grands nombres, on a

$$\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \xrightarrow{P} \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t).$$

Maintenant, en appliquant le théorème central limite, nous avons

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{x}_i y_i - \mathbf{x}_i \mathbf{x}_i^t \beta) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \epsilon_i \xrightarrow{d} \mathcal{N}(0, \sigma^2 \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t))$$

où, $\Sigma = \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t)$. Le résultat du théorème découle alors du théorème de Slutsky.

Malgré l'efficacité et la simplicité de l'estimation par la méthode des MCO, cette approche souffre de problème de robustesse aux points aberrants. D'où la nécessité d'améliorer cette méthode. Dans les prochaines sections, nous présentons certaines approches plus robustes.

En présence de plusieurs variables explicatives, le choix des variables qui contribue significativement à la modélisation de la variable réponse est très important. On cherche un modèle avec juste les variables importantes. Lorsque p est petit, souvent, on utilise les méthodes *Stepwise*, AIC et BIC. Or, lorsque le nombre de variables explicatives est très important, ces méthodes sont inutiles et sans efficacité.

1.2 La méthode Lasso

Dans cette partie, on s'intéresse au cas où le nombre des variables explicatives est plus grand que le nombre d'observations, c-à-d quand $n \ll p$. Dans ce cas, la matrice $\mathbf{X}^t \mathbf{X}$ n'est plus inversible. L'idée de la méthode lasso (*least absolute shrinkage and selection operator*) est donc, de faire une pénalisation de norme L_1 qui force certains coefficients à 0. voir [HT15].

Définition 1.7

L'approche de la régression LASSO minimise le critère suivant :

$$\|Y - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1, \quad \lambda \geq 0$$

ce qui revient à résoudre le problème, en fixant une valeur $t \geq 0$,

$$\begin{cases} \min_{\boldsymbol{\beta}} \|Y - \mathbf{X}\boldsymbol{\beta}\|_2^2 \\ S.C. \quad \|\boldsymbol{\beta}\|_1 \leq t. \end{cases}$$

La méthode LASSO est utilisée pour contrôler la complexité de $\boldsymbol{\beta}$, ce qui :

- force certains coefficients, non importants, à 0.
- favorise des modèles parcimonieux, puisqu'une sélection de variables est automatiquement faite.

Solution analytique

Pour un paramètre de régularisation $\lambda \in \mathbb{R}_+$, on peut distinguer 3 cas possibles :

1. $\lambda = 0$, on retrouve ainsi la solution des moindres carrés.
2. $\lambda \in]0, \infty[$, implique que l'augmentation du paramètre λ conduit à la diminution de certains coefficients $\boldsymbol{\beta}$ vers 0 jusqu'à ce qu'ils soient nuls. La solution est appelée parcimonieuse (creuse) car elle comporte plusieurs coefficients nuls.

3. $\lambda \longrightarrow \infty$ pénaliser très fortement tous les coefficients non-égaux à zéro, ce qui finalement, force toutes les composantes de β à 0

Il y a donc ici un compromis à faire entre l'ajustement des données et la pénalité.

1.2.1 Interprétation géométrique

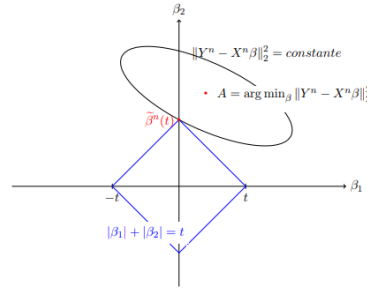


Figure 1.1 – Cas où la solution se trouve dans la zone qui annule le coefficient β_2 .

La figure 1.1 montre que la solution, $\tilde{\beta}^n(t)$, par la méthode LASSO est un point situé à l'intersection d'une ligne de niveau du terme d'erreur et de la région $\|\beta\|_1 \leq t$ de \mathbb{R}^d où la contrainte est vérifiée.

1.2.2 Régression linéaire robuste

Dans cette partie, on doit premièrement introduire la notion de robustesse et celle des points aberrants.

Robustesse

La robustesse en statistique est définie comme la capacité d'un modèle à maintenir ses performances malgré la variabilité de ses paramètres. Généralement, on ne cherche pas

à éliminer les causes d'une variabilité, mais plutôt à minimiser ses effets. Généralement, dans la modélisation, on cherche un modèle statistique avec un minimum d'hypothèses. Les méthodes robustes garantissent de bons résultats même si certaines des hypothèses utilisées pour justifier la méthode d'estimation ne sont pas applicables.

Points aberrants

Point aberrant (outlier en anglais) est une observation ou un sous-ensemble d'observations, qui paraissent être différents en comparant avec la majorité de l'ensemble des données. On peut distinguer trois différents types de points aberrants (voir figure 1.2) :

- **Point vertical** (suivant l'axe des Y) : ce sont des valeurs atypiques pour la variable Y à expliquer. Ils ont une influence relativement faible, ils perturbent généralement le biais plutôt que la pente de la droite de la régression.
- **Point horizontal** (suivant l'axe des X) : dit point levier, il s'agit d'une ou plusieurs valeurs atypiques sur l'abscisse des variables explicatives. Il y a deux sortes de points horizontaux :
 - **Bon point levier** : ce sont les points avec un bon effet, c'est-à-dire qu'ils ont une valeur résiduelle du même ordre que celle de la majorité des données.
 - **Mauvais point levier** : ce sont les points qui ont un effet défavorable, car ils entraînent un changement de la pente.

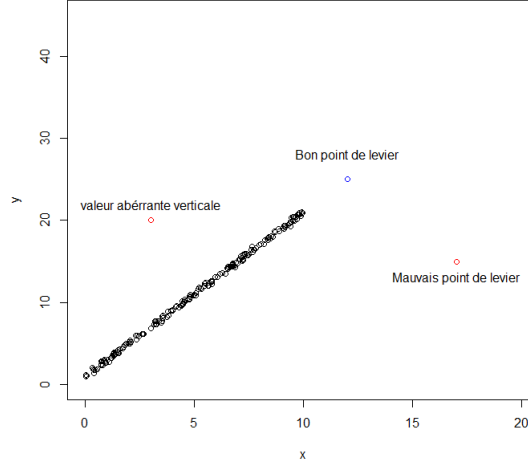


Figure 1.2 – Différents points aberrants.

Maintenant, à partir des points aberrants on définit le concept du point de rupture qui permet de mesurer la robustesse des méthodes d'estimation. En fait, c'est le pourcentage de valeurs aberrantes qui mène à une estimation biaisée des paramètres du modèle. Pour plus de détails, on recommande le livre de Donoho et Huber [DH83] et l'article de Yohai [Yoh87].

Définition 1.8 Soient $\hat{\theta}(\mathcal{X})$ un estimateur basé sur l'échantillon \mathcal{X} et $\hat{\theta}(\mathcal{X}_m^*)$ l'estimateur basé sur l'échantillon \mathcal{X}_m^* obtenu en remplaçant m valeurs arbitraires dans l'échantillon initial, \mathcal{X} . Le point de rupture pour $\hat{\theta}$ est la plus petite fraction de valeurs arbitraires que l'on doit remplacer dans l'échantillon \mathcal{X} pour obtenir un écart infini entre $\hat{\theta}(\mathcal{X})$ et $\hat{\theta}(\mathcal{X}_m^*)$:

$$\epsilon^*(\hat{\theta}) = \frac{1}{n} \inf \left\{ m \left| \sup_{\mathcal{X}_m^*} \|\hat{\theta}(\mathcal{X}) - \hat{\theta}(\mathcal{X}_m^*)\|_2 = \infty \right. \right\}.$$

Par exemple, l'estimation des moindres carrés à un point de rupture de $\frac{1}{n}$, qui tend vers 0 si n tend vers l'infini, car il suffit d'un seul point aberrant pour obtenir une mauvaise

estimation des paramètres. La figure 1.3 montre l'effet d'un seul point de levier sur l'estimation par les MCO. Plus ce point s'éloigne du nuage des points, plus l'écart croît entre les estimateurs basés sur l'échantillon sans et avec ce point.

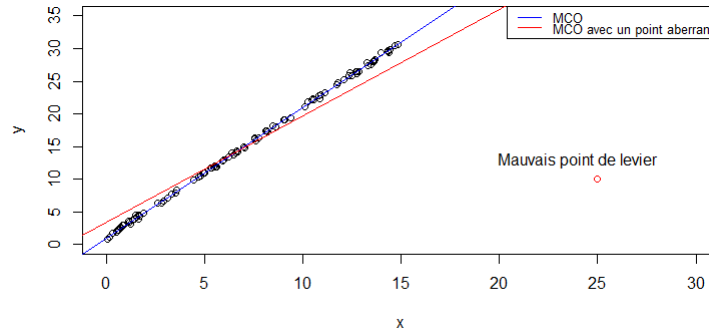


Figure 1.3 – Estimateur par MCO avec des données aberrantes.

Estimateur robuste de la régression linéaire

La sensibilité de la méthode des MCO provient du poids important attribué aux valeurs aberrantes. Donc, une solution naturelle consiste à diminuer ce poids et de considérer des fonctions de perte qui croissent de façon quadratique pour les petites valeurs des erreurs, mais qui croissent de façon moins rapide ou pas du tout quand on les erreurs s'éloignent de zéro. Ceci permet de construire des estimateurs plus robustes aux points aberrants.

Parmi les estimateurs robustes les plus fréquents, on trouve le M-estimateur, le S-estimateur et le MM-estimateur. Dans ce travail, nous présentons le M-estimateur, voir [Hub64] et [Hub81]. Cette approche est une généralisation de l'estimateur du maximum de vraisemblance.

On considère le modèle (1.2) et on définit la fonction ψ dite "fonction d'influence" ou ψ -fonction. Une ψ -fonction est une fonction continue par morceaux définie de \mathbb{R} dans \mathbb{R} telle que

- (1) ψ est symétrique par rapport à l'origine ($\psi(-x) = -\psi(x)$ pour tout x dans \mathbb{R}).
- (2) $\psi(x) \leq 0$ pour $x \leq 0$ et $\psi(x) > 0$ pour $0 < x < x_r$, où $x_r = \sup\{x : \psi(x) > 0\}$.

Maintenant on introduit la fonction de perte ou ρ -fonction qui est la fonction primitive de ψ :

$$\rho(x) = \int_0^x \psi(t)dt.$$

On peut facilement déduire que :

- (1) ρ est non-négative,
- (2) $\rho(0) = 0$, ρ est une fonction symétrique ($\rho(-x) = \rho(x)$),
- (3) $\frac{d\rho(x)}{dx} = \psi(x)$,
- (4) ρ monotone en $|\epsilon_i|$, $\rho(\epsilon_i) \geq \rho(\epsilon_{i'})$, pour $|\epsilon_i| > |\epsilon_{i'}|$.

Les choix des fonctions de perte ρ les plus populaires sont celles proposées par Huber ([Hub64]) et Tukey ([Hub81]). La table 1.1 présente ces deux fonctions objectifs. La valeur k des estimateurs de Huber et Tukey, appelée " constante de réglage (tuning constant)". La figure 1.4 représente les fonctions objectifs dans la table 1.1. On voit que les trois fonctions ont un comportement similaire pour les résidus proches du zéro, mais plus le résidu s'éloigne de zéro, plus la méthode des MCO donne un poids important aux valeurs, alors que les autres approches croissent lentement (Huber) ou pas du tout (Tukey).

TABLE 1.1 – Exemples des fonctions objectifs utilisées pour les M-estimateurs, $k \geq 0$

Méthode	Fonction objectif
Moindres-carrés	$\rho_{MC}(\epsilon) = \epsilon^2$
Huber	$\rho_H(\epsilon) = \begin{cases} \frac{1}{2}\epsilon^2, & \epsilon \leq k \\ k \epsilon - \frac{1}{2}k^2, & \epsilon > k. \end{cases}$
Tukey	$\rho_T(\epsilon) = \begin{cases} \frac{k^2}{6}\{1 - [1 - (\frac{\epsilon}{k})^2]^3\}, & \epsilon \leq k \\ k^2/6, & \epsilon > k. \end{cases}$

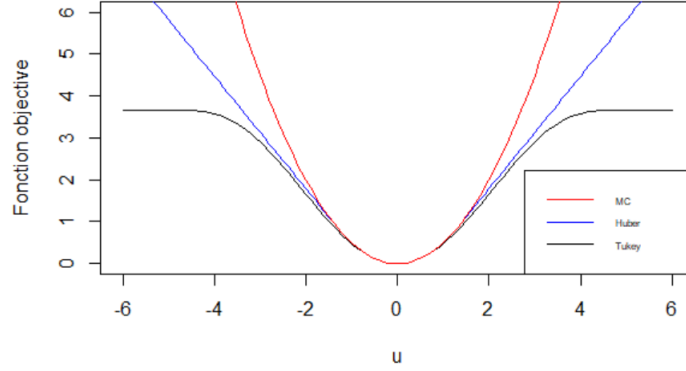


Figure 1.4 – Fonctions objectives pour MC, Huber ($k = 1, 345$) et Tukey ($k = 4, 685$).

L'estimateur par la méthode M-estimation est donné par

$$\begin{aligned}\hat{\beta}^M &= \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n \rho(\epsilon_i) \\ &= \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n \rho(y_i - (\beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p})).\end{aligned}$$

Le coefficient de détermination, associé à l'estimateur des MCO, qui mesure la qualité du modèle n'est pas sensible aux changements d'unité de la variable dépendante. La méthode M-estimation ne possède pas cette propriété. Pour résoudre ce problème, on définit par $\hat{\epsilon}_i := y_i - \hat{y}_i$, le résidu qui est l'écart entre l'observation y_i et sa prévision $\hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1^M x_{i,1} + \dots + \hat{\beta}_p^M x_{i,p}$. Pour standardiser les résidus, on introduit un estimateur robuste pour la dispersion des résidus, [And08] :

$$\hat{\sigma} = 1,4826 MAD \tag{1.3}$$

où MAD signifie "Median absolute deviation" c-à-d $MAD = \text{médiane } |\hat{\epsilon}_i|$. En supposant la normalité des erreurs ($N(0, \sigma^2)$), la valeur 1,4826 provient du fait que

$$0,5 \approx \mathbb{P} \left(\left| \frac{\epsilon_i}{\sigma} \right| \leq \frac{MAD}{\sigma} \right)$$

et donc, 1.4826 est le quantile de la distribution normale absolue évalué à 0,5.

Le M-estimateur invariant par changement d'échelle est alors défini par

$$\hat{\beta}^M = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^n \rho((y_i - (\beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p}))/\hat{\sigma}).$$

Cet estimateur est obtenu en résolvant un système de $p + 1$ équations suivant :

$$\begin{cases} \sum_{i=1}^n \frac{\partial \rho(\hat{\epsilon}_i/\hat{\sigma})}{\partial \beta_0} = \sum_{i=1}^n \psi(\hat{\epsilon}_i/\hat{\sigma}) & = 0 \\ \sum_{i=1}^n \frac{\partial \rho(\hat{\epsilon}_i/\hat{\sigma})}{\partial \beta_j} = \sum_{i=1}^n \psi(\hat{\epsilon}_i/\hat{\sigma}) x_{ij} & = 0, \quad \forall j = 1, \dots, p. \end{cases}$$

Pour résoudre le système d'équations (1.4), [DS98] ont proposé une réécriture de (1.4) en posant $\omega_i = \frac{\psi(\hat{\epsilon}_i/\hat{\sigma})}{\hat{\epsilon}_i/\hat{\sigma}}$.

En rappelant que $\hat{\epsilon}_i = y_i - (\hat{\beta}_0^M + \hat{\beta}_1^M x_{i,1} + \dots + \hat{\beta}_p^M x_{i,p})$, ce système d'équations (1.4) s'écrit en fonction des poids ω_i comme suit :

$$\begin{cases} \sum_{i=1}^n (y_i - (\hat{\beta}_1^M + \hat{\beta}_1^M x_{i,1} + \dots + \hat{\beta}_p^M x_{i,p})) \omega_i & = 0 \\ \sum_{i=1}^n (y_i - (\hat{\beta}_1^M + \hat{\beta}_1^M x_{i,1} + \dots + \hat{\beta}_p^M x_{i,p})) x_{ij} \omega_i & = 0, \quad \forall j = 1, \dots, p. \end{cases}$$

Cette dernière peut s'écrire sous la forme matricielle suivante :

$$\mathbf{X}^t \mathbf{W} \mathbf{X} \hat{\beta}^M = \mathbf{X}^t \mathbf{W} \mathbf{Y}$$

où la matrice $\mathbf{W} = \operatorname{diag}(w_1, w_2, \dots, w_n)$ est une matrice diagonale $n \times n$ contenant tous les poids w_i . Si la matrice $\mathbf{X}^t \mathbf{W} \mathbf{X}$ est inversible, alors le M-estimateur s'écrit

$$\hat{\beta}^M = (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} \mathbf{Y}.$$

Comme la fonction de poids ω_i dépend des résidus, lesquels dépendent des coefficients de la régression, une solution itérative (IRLS) est nécessaire pour résoudre ce système. Susanti, Pratiwi, Sulistijowati et Liana [SPHL13] ont proposé un algorithme pour résoudre

ces systèmes d'équations.

Sous certaines hypothèses, [JSP13] (pages 215 à 224) ont montré la convergence en loi du M-estimateur (voir aussi [Yoh87]).

Théorème 1.9 *Sous certaines conditions, voir [JSP13] ,*

$$\sqrt{n}\Sigma^{1/2}(\hat{\beta}^M - \beta) \xrightarrow{d} N(0, \sigma^2 \mathbb{I}_{(p+1)(p+1)}/e)$$

où Σ est définie dans le théorème 1.6 et où $e = \frac{[\int \psi'(x)\phi(x)dx]^2}{\int \psi^2(x)\phi(x)dx}$ où ϕ désigne la densité d'une loi normale centrée et réduite.

Remarque 1.10 *Voici quelques remarques concernant les M-estimateurs :*

1. La constante e détermine l'efficacité relative de $\hat{\beta}^M$ par rapport à $\hat{\beta}^{MCO}$ donnée par $\frac{\text{Var}(\hat{\beta}^{MCO})}{\text{Var}(\hat{\beta}^M)}$. Les valeurs de k dans la figure 1.4 sont déterminées en posant $e = 0,95$.
2. Le M-estimateur est robuste aux points verticaux, mais pas aux points de levier. Comme la méthode des MCO, le point de rupture de M-estimation est égal à $1/n$, qui tend vers zéro si n est assez grand, voir [RL87].
3. le S-estimateur est proposé, par [DSTA15], pour obtenir un estimateur plus robuste avec un grand point de rupture. Cependant, cet estimateur souffre de problème de l'inefficacité relative comparé à l'estimateur des MCO, voir [RY84].
4. Pour obtenir un estimateur robuste ayant une bonne efficacité relative, le MM-estimateur est construit. Cet estimateur est une combinaison de M-estimateur et S-estimateur, voir [SPHL13]. Dans le chapitre 4, nous proposons un nouvel estimateur facile à implémenter, robuste avec une efficacité proche de celle des estimateurs des MCO.

1.2.3 Régression linéaire médiane (LAD)

Dans cette section, nous présentons un estimateur robuste aux points aberrants, appelé régression médiane. Récemment, avec le développement des méthodes et algorithmes d'optimisation, cette méthode est devenue plus populaire que la régression moyenne. D'abord, nous commençons par présenter la médiane non-conditionnelle.

Soit Y une variable aléatoire de fonction de répartition F_Y , la médiane de Y est donnée par :

$$Med(y) = F_Y^{-1}(\frac{1}{2})$$

où F_Y^{-1} , dite aussi fonction quantile, est l'inverse généralisé de F_Y donnée par

$$F_Y^{-1}(y) = \inf_x \{F_Y(x) \geq y\}.$$

On peut définir également la médiane comme suit :

$$Med(y) = \operatorname{argmin}_c E[|Y - c| - |Y|]$$

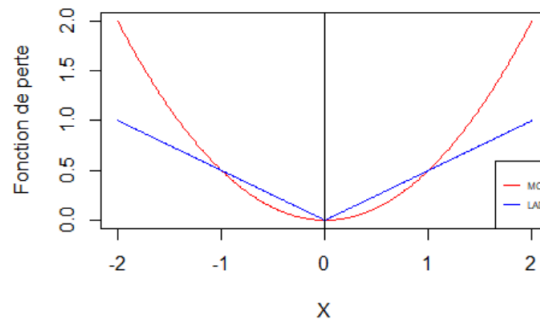


Figure 1.5 – Fonctions de perte associées à la moyenne et la médiane.

La figure 1.5 illustre les fonctions de perte utilisées pour trouver la médiane et la moyenne. On peut remarquer que la perte absolue augmente de façon linéaire donc moins rapide

que la perte quadratique. En effet, la médiane ne donne pas un poids très élevé aux grandes observations. Cet estimateur est connu comme un indicateur central robuste de la variable aléatoire.

Régression médiane

On considère le modèle 1.2 et on suppose que $\epsilon_1, \dots, \epsilon_n$ sont i.i.d. de médiane égale à zéro. Au lieu d'utiliser le carré comme mesure de perte, la régression par rapport à la médiane (régression L_1) utilise la valeur absolue :

$$\hat{\beta}^L = \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n |y_i - \mathbf{x}_i^t \beta|. \quad (1.4)$$

Cette méthode appelée "Least absolute deviation (LAD)", est utilisée comme une alternative robuste à l'estimation des MCO. Elle est équivalente à la méthode de maximum de vraisemblance avec la distribution de Laplace pour les erreurs, au lieu de la normale pour la MCO.

Contrairement à l'estimation par MCO et comme la norme L_1 n'est pas dérivable partout (au point minimum x_0), la solution de l'équation (1.4) n'est pas explicite. Une solution numérique est suggérée. Une des méthodes les plus utilisées est celle basée sur la programmation linéaire avec contrainte. En effet, le problème (1.4) peut s'écrire comme un programme linéaire suivant [Koe05]

$$\min_{(\beta, \mathbf{u}, \mathbf{v}) \in (\mathbb{R}^{p+1} \times \mathbb{R}^{2n})} \{\tau \mathbf{1}_n^t \mathbf{u} + (1 - \tau) \mathbf{1}_n^t \mathbf{v} \mid \mathbf{X}\beta + \mathbf{u} + \mathbf{v} = \mathbf{Y}\},$$

où $\mathbf{1}_n$ est un vecteur de 1 de taille n et $\tau = 0, 5$. Plusieurs algorithmes ont été proposés pour résoudre ce programme, parmi ceux-ci on trouve l'algorithme du simplexe et points intérieurs.

Remarque 1.11 *La figure 1.6 montre une illustration de la régression médiane en présence d'un point de levier. On remarque que cette régression suit les nuages des points et*

n'est pas affectée par la présence des valeurs aberrantes contrairement à la méthode des MCO et M-estimation. En fait, la régression médiane est basé sur les rangs et non sur les valeurs des observations.

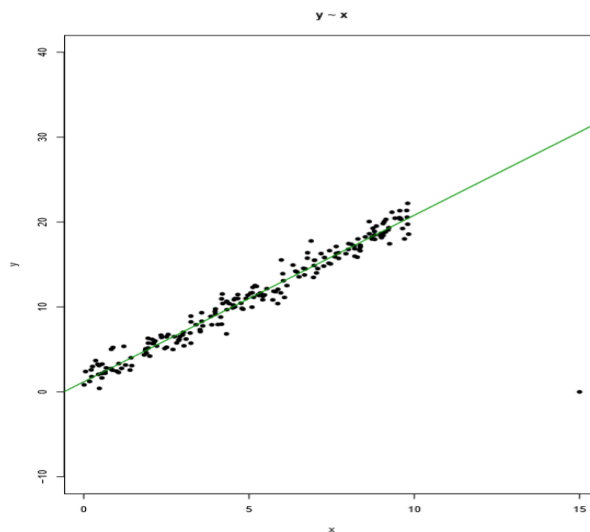


Figure 1.6 – La régression médiane avec un point de levier.

Le théorème suivant établit la normalité asymptotique de l'estimateur $\hat{\beta}^L$

Théorème 1.12 *Soit $f_{\epsilon|\mathbb{X}}$ est la densité de ϵ conditionnelle à \mathbb{X} .*

La distribution asymptotique de l'estimateur $\hat{\beta}^L$ est donnée par :

$$\sqrt{n}\Sigma^{1/2}(\hat{\beta}^L - \beta) \sim N\left(0, \frac{1}{4f_{\epsilon|\mathbb{X}}^2(0)}\mathbb{I}_{(p+1)(p+1)}\right)$$

où Σ définie au théorème 1.6.

La preuve de ce résultat se trouve dans [Koe05], pages 120 à 122. On voit bien que la variance de $\hat{\beta}^L$ dépend fortement de la valeur de la densité conditionnelle en zéro. Plus la concentration des erreurs est forte en zéro, plus l'efficacité de cet estimateur est grande.

1.3 Régression linéaire basée sur les pertes asymétriques

Dans la section précédente, nous avons étudié la tendance centrale de la variable d'intérêt Y conditionnelle aux variables explicatives \mathbb{X} . Ces régressions ne donnent pas toutes les informations sur le comportement de Y conditionnelle à \mathbb{X} . Pour cela, on a besoin de modéliser la distribution conditionnelle de Y étant donné \mathbb{X} . Souvent la modélisation de cette distribution est assez compliquée et nous avons besoin d'un autre outil pour caractériser le comportement de Y conditionnelle à \mathbb{X} . Dans cette section, nous allons présenter deux modèles de régression qui peuvent caractériser la distribution conditionnelle de Y étant donné \mathbb{X} : La régression quantile et expectile. Ces deux régressions sont des généralisations de la régression moyenne et médiane. L'idée derrière ces régressions est d'utiliser des fonctions de perte asymétriques quand on s'éloigne du centre. Nous commençons par l'introduction des quantiles et expectiles non-conditionnelles et établir le lien entre ces deux paramètres. Ensuite, nous allons définir et étudier les propriétés asymptotiques des deux régressions : quantile et expectile.

1.4 Fonctions de perte asymétriques

Dans cette section, nous introduisons les fonctions de perte asymétriques, quantiles et expectiles d'une variable aléatoire et la relation entre elles.

1.4.1 Quantiles

Les quantiles permettent de caractériser la distribution d'une variable. En effet, ces paramètres donnent toutes les informations sur la concentration d'une variable. Notons que

la médiane est un cas particulier des quantiles.

Définition 1.13 Soit Y une variable aléatoire, de fonction de répartition F_Y ($F_Y(y) = P(Y \leq y)$). Le quantile d'ordre $\tau \in (0, 1)$ est donné par :

$$q_\tau(Y) = \inf_y \{y : F_Y(y) \geq \tau\}.$$

On note souvent $q_\tau(Y)$ par $F_Y^{-1}(\tau)$ et appelé inverse généralisé de F_Y . Les quantiles les plus utilisés sont : le premier quartile $q_{0,25}(Y)$, la médiane $q_{0,5}(Y)$ et le troisième quartile $q_{0,75}(Y)$. Remarquons aussi que les moments d'ordre k de la variable Y peuvent s'écrire, en utilisant un changement de variables $\tau = F_Y(y)$, de la façon suivante :

$$\mathbb{E}(Y^k) = \int_{-\infty}^{+\infty} y^k dF_Y(y) = \int_0^1 [q_\tau(Y)]^k d\tau.$$

On peut définir également le quantile $q_\theta(Y)$ comme solution du problème de minimisation suivant :

$$\operatorname{argmin}_{\theta \in \mathbb{R}} \mathbb{E}[\rho_\tau(Y - \theta)],$$

où la fonction de perte $\rho_\tau(\cdot)$ est définie par :

$$\rho_\tau(u) = \begin{cases} \tau |u|, & u \geq 0 \\ (1 - \tau) |u|, & u < 0. \end{cases} \quad (1.5)$$

En effet,

$$\operatorname{argmin}_{\theta \in \mathbb{R}} \mathbb{E}[\rho_\tau(Y - \theta)] = \operatorname{argmin}_{\theta \in \mathbb{R}} \left[(\tau - 1) \int_{-\infty}^{\theta} (y - \theta) dF_Y(y) + \tau \int_{\theta}^{\infty} (y - \theta) dF_Y(y) \right]$$

En dérivant par rapport θ , le minimum q_τ satisfait

$$\begin{aligned}
0 &= (\tau - 1) \left\{ [(y - \theta)f(y)]_{y=-\infty}^{\theta} + \int_{-\infty}^{\theta} \frac{\partial}{\partial \theta}(y - \theta) dF_Y(y) \right\} \\
&+ \tau \left\{ [(y - \theta)f(y)]_{y=\theta}^{\infty} + \int_{\theta}^{\infty} \frac{\partial}{\partial \theta}(y - \theta) dF_Y(y) \right\} \\
&= (\tau - 1) \{0 - F_Y(\theta)\} + \tau \{0 - (1 - F(\theta))\} \\
&= F_Y(\theta) - \tau
\end{aligned}$$

Puisque $F_Y(\theta)$ est croissante, donc sa dérivée est positive. Par conséquent, la valeur θ telle que $F_Y(\theta) - \tau = 0$ est un minimum.

Ce qui implique :

$$\tau = F_Y(q_{\tau}).$$

La figure 1.7 illustre les pertes ρ_{τ} pour trois valeurs de τ . La valeur $\tau = 0,5$ correspond à la médiane et donne des poids symétriques autour de zéro. Le cas $\tau = 0,75$, les observations à droite reçoivent un poids élevé contrairement à celles qui sont à gauche. Par contre, si $\tau = 0,25$, les observations à gauche, les petites valeurs de Y , reçoivent un poids élevé et celles de gauche reçoivent un poids plus faible.

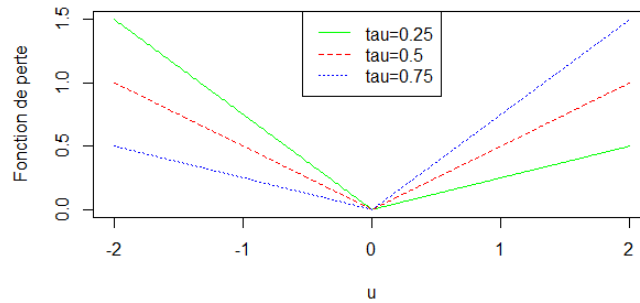


Figure 1.7 – Fonction de perte des quantiles.

1.4.2 Expectiles

L'expectile est un concept introduit par Newey et Powell [NP87] et moins étudié comparé au quantile. Si le quantile est une généralisation de la médiane, l'expectile est une généralisation de la moyenne. Aussi, le quantile est basé sur la fonction de perte absolue et l'expectile utilise la fonction de perte quadratique.

Définition 1.14 *L'expectile de niveau $\omega \in (0, 1)$ d'une variable aléatoire Y est défini par :*

$$\mu_\omega(Y) = \underset{\theta \in \mathbb{R}}{\operatorname{argmin}} \mathbb{E}[r_\omega(Y - \theta)]$$

où $r_\omega(\cdot)$ est défini par :

$$r_\omega(u) = \begin{cases} \omega u^2, & u \geq 0 \\ (1 - \omega) u^2, & u < 0 \end{cases} \quad (1.6)$$

et appelé la fonction de perte asymétrique des moindres carrés.

La figure 1.8 présente la fonction de perte des expectiles. Dans le cas $\omega = 0,5$, on a la fonction de perte pour retrouver la moyenne. Le cas où $\omega = 0,75$, on voit qu'on pénalise beaucoup à droite par rapport au gauche, en donnant un poids plus élevé et le contraire quand $\omega = 0,25$.

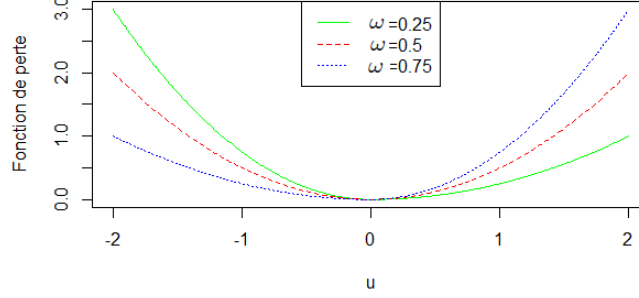


Figure 1.8 – Fonction de perte des expectiles.

On utilise les conditions d’optimalité de premier ordre (*KKT1* *abréviation de Karush-Kuhn-Tucker*) sur,

$$\mathbb{E}[r_\omega(Y - \mu)] = (1 - \omega) \int_{(-\infty, \mu]} (y - \mu)^2 dF(y) + \omega \int_{[\mu, \infty)} (y - \mu)^2 dF(y).$$

Donc, en dérivant par rapport à μ et en égalant la dérivée à 0, on trouve :

$$0 = (1 - \omega) \int_{-\infty}^{\mu_\omega(Y)} (y - \mu) dF(y) + \omega \int_{\mu_\omega(Y)}^{\infty} (y - \mu) dF(y). \quad (1.7)$$

[NP87] ont montré que l’expectile $\mu_\omega(Y)$ est la solution unique si $\mathbb{E}(Y) = \mu_{0,5}(Y)$ existe. Cette solution s’écrit comme suit :

$$\mu_\omega(Y) = \mathbb{E}(Y) - \frac{1 - 2\omega}{1 - \omega} \mathbb{E}[\{Y - \mu_\omega(Y)\} \mathbb{1}\{Y > \mu_\omega(Y)\}]. \quad (1.8)$$

Malheureusement, cette formule ne peut être utilisée pour calculer directement $\mu_\omega(Y)$ pour toutes les valeurs de ω .

Remarque : La moyenne est l’expectile le plus populaire. Il suffit de choisir $\omega = 0, 5$, on trouve que $\mu_{0,5} = \mathbb{E}(Y)$. Pour plus de propriétés sur l’expectile, voir [HBK16]

1.4.3 Relation entre les quantiles et les expectiles

À cause de la similarité des propriétés des quantiles et des expectiles, plusieurs études ont été entreprises pour étudier la relation entre ces deux statistiques. Parmi les chercheurs qui ont exploré cette relation, on trouve [Jon94]. En fait, les quantiles et les expectiles sont des statistiques équivalentes pour des distributions absolument continues, à chaque quantile q_τ , de niveau τ , on peut trouver un expectile $\mu_\omega(Y)$, de niveau ω , tel que $q_\tau = \mu_\omega(Y)$. En effet, fixons une valeur de τ et soit q_τ le quantile correspondant. Si on cherche la valeur de ω telle que son expectile $\mu_\omega(Y)$ soit égal à q_τ , il suffit de remplacer q_τ par $\mu_\omega(Y)$ dans l'équation (1.7). On trouve alors :

$$\omega = \frac{\int_{-\infty}^{q_\tau} (|y - q_\tau|) f_Y(y) dy}{\int_{-\infty}^{+\infty} (|y - q_\tau|) f_Y(y) dy}.$$

Aussi, Abdous et Rémillard [AR95] ont montré que la distribution F_Y des quantiles de Y peut s'écrire en fonction de la distribution G des expectiles. En effet, le fait que $\omega = G(\mu_\omega(Y))$ et d'après (1.7)

$$\omega = \frac{\int_{-\infty}^{\mu_\omega(Y)} (|y - \mu|) f_Y(y) dy}{\int_{-\infty}^{+\infty} (|y - \mu|) f_Y(y) dy}$$

on a

$$G(t) = \frac{-tF(t) + P(t)}{\mu - 2P(t) - t(1 - 2F(t))}$$

où $P(t) = \int_{-\infty}^t y f_Y(y) dy$ et où $\mu = \int_{-\infty}^{+\infty} y f_Y(y) dy$.

La densité correspondante à cette distribution est donnée par :

$$g(t) = \frac{\mu F(t) - P(t)}{[\mu - 2P(t) - t(1 - 2F(t))]^2}$$

La figure 1.9 présente les expectiles et les quantiles de la distribution normale. on peut voir que les courbes de cette dernière se croisent quand $\tau = \omega = 0,5$.

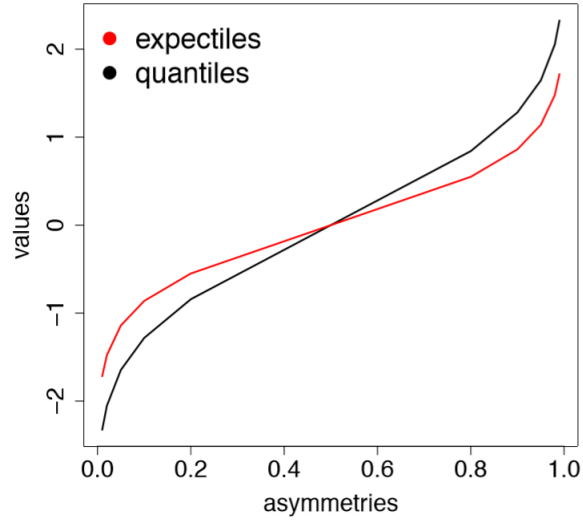


Figure 1.9 – Expectiles et quantiles de la distribution normale.

1.5 Régression quantile et régression expectile

Dans le cadre classique de la régression, on souhaite décrire la relation entre la variable d'intérêt (réponse) et les variables explicatives (régresseurs), en estimant spécifiquement l'effet de ces dernières sur la médiane ou la moyenne de la distribution de la variable réponse. Dans cette partie, nous introduisons des modélisations plus générales qui permettent une analyse plus détaillée et globale de l'influence des régresseurs sur la variable d'intérêt, d'où la régression quantile (RQ) et la régression expectile (RE).

1.5.1 Régression quantile linéaire

Dans la section précédente, nous avons introduit les quantiles non-conditionnels, c-à-d sans l'effet des autres variables. Cette définition peut être généralisée au cas conditionnel. En présence d'un vecteur de covariables \mathbb{X} , on cherche le quantile d'une variable réponse Y en fonction de la valeur de ces covariable, c-à-d, pour un quantile τ , le quantile condi-

tionnel de Y sachant \mathbb{X} est donné par

$$q_\tau(Y|\mathbb{X}) = \inf \{y : F_{Y|\mathbb{X}}(y) \geq \tau\}.$$

où $F_{Y|(X_1, \dots, X_d)}$ est la fonction de répartition de Y sachant \mathbb{X} .

Comme le quantile non-conditionnel, $q_\tau(Y|\mathbb{X})$ peut s'écrire sous la forme d'une minimisation de la fonction de perte

$$\operatorname{argmin}_{\theta \in \mathbb{R}} \mathbb{E} [\rho_\tau(Y - \theta) | \mathbb{X}]$$

où ρ_τ est défini par (1.5).

Dans ce travail, nous considérons le modèle de régression quantile linéaire défini comme suit :

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}_\tau + \boldsymbol{\epsilon}_\tau, \text{ tel que } q_\tau(\boldsymbol{\epsilon}_{i,\tau} | \mathbf{x}_i^t) = 0.$$

où pour chaque τ correspond un vecteur de coefficient $\boldsymbol{\beta}_\tau = (\beta_0^\tau, \beta_1^\tau, \dots, \beta_p^\tau)$. Cette dernière expression implique que :

$$q_\tau(Y_i | \mathbf{x}_i^t) = \mathbf{x}_i^t \boldsymbol{\beta}_\tau.$$

Estimation

L'estimation du paramètre $\boldsymbol{\beta}_\tau$ du modèle de régression quantile est obtenue en minimisant le critère suivant, voir [KB78]

$$\hat{\boldsymbol{\beta}}_\tau = \operatorname{argmin}_{\boldsymbol{\beta} \in \mathbb{R}} \sum_{i=1}^n \rho_\tau(y_i - \beta_0^\tau - \beta_1^\tau x_{i,1} - \dots - \beta_p^\tau x_{i,p}).$$

où $\rho_\tau(u) = u\psi_\tau(u)$, avec $\psi_\tau(u) = (\tau - \mathbb{1}(u < 0))$, est la fonction de perte associée à la RQ

Une solution explicite de cette minimisation n'existe pas. À partir des différents algorithmes, tel que l'algorithme du simplexe et point intérieur, on peut générer les solutions de ce dernier problème d'optimisation (voir [CW05]). La figure 1.10 illustre la régression quantile linéaire pour différentes valeurs de τ .

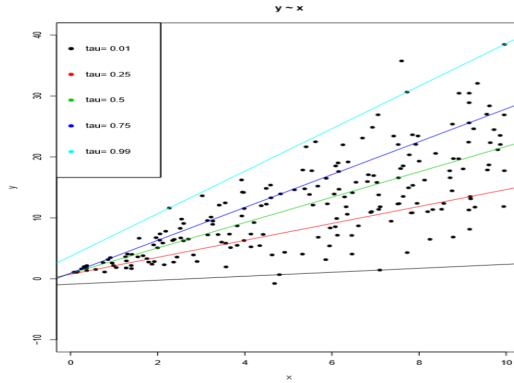


Figure 1.10 – La regression quantile pour différents τ .

L'estimation des paramètres du modèle de régression linéaire quantile est la minimisation des erreurs avec la fonction de perte ρ_τ . Ceci est équivalent à la maximisation de la vraisemblance en supposant la loi de Laplace asymétrique à la variable de réponse donnée par

$$f(y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp \left(-\rho_\tau \left(\frac{y-\mu}{\sigma} \right) \right).$$

Donc, en posant $\mu_i = \mathbf{x}_i^t \boldsymbol{\beta}$, la vraisemblance est de la forme suivante

$$\mathbb{L}(\boldsymbol{\beta}, \sigma | \mathbf{y}, \tau) \propto \sigma^{-n} \prod_{i=1}^n \exp \left(-\rho_\tau \left(\frac{y_i - \mathbf{x}_i^t \boldsymbol{\beta}}{\sigma} \right) \right).$$

Le théorème suivant établit la normalité asymptotique de l'estimateur $\hat{\boldsymbol{\beta}}_\tau$. Les hypothèses ne sont pas citées dans ce travail. Pour plus de détails, voir le livre [Koe05] ou l'article [PK97].

Théorème 1.15 (Normalité asymptotique de $\widehat{\beta}_\tau$) *Sous certaines conditions de régularité, la loi asymptotique de $\widehat{\beta}_\tau$ est donnée par, voir [PK97] :*

$$\sqrt{n}\Sigma^{1/2}(\widehat{\beta}_\tau - \beta_\tau) \xrightarrow{d} N\left(0, \frac{\tau(1-\tau)}{f_{\epsilon_\tau|\mathbb{X}}^2(0)}\mathbb{I}_{(p+1)(p+1)}\right), \quad (1.9)$$

où Σ est définie au théorème 1.6.

Démonstration

On définit pour $\delta \in \mathbb{R}^{p+1}$

$$A_n(\delta) = \sum_{i=1}^n \left[\rho_\tau \left(\epsilon_{i,\tau} - \frac{1}{\sqrt{n}} \delta^t \mathbf{x}_i \right) - \rho_\tau(\epsilon_{i,\tau}) \right]$$

où $\rho_\tau(u) = (\tau - \mathbb{1}(u \leq 0))u = \psi_\tau(u)u$, $\epsilon_{i,\tau} = y_i - \mathbf{x}_i^t \beta_\tau$. Notons que la dérivée de $\rho_\tau(u)$ est égale à $\psi(u)$.

On peut montrer que :

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + R(u, v),$$

où $R(u, v) = [\mathbb{1}(u \leq 0) - \mathbb{1}(u \leq v)](u - v)$.

En effet,

$$\begin{aligned} \rho_\tau(u - v) - \rho_\tau(u) &= (\tau - \mathbb{1}(u \leq v))(u - v) - (\tau - \mathbb{1}(u \leq 0))u \\ &= (\tau - \mathbb{1}(u \leq v))(u - v) - (\tau - \mathbb{1}(u \leq 0))(u - v) - (\tau - \mathbb{1}(u \leq 0))v \\ &= -v\psi_\tau(u) + R(u, v). \end{aligned}$$

D'après cette décomposition et en considérant $u = \epsilon_{i,\tau}$ et $v = \frac{1}{\sqrt{n}} \delta^t \mathbf{x}_i$, $A_n(\delta)$ peut s'écrire comme suit :

$$A_n(\delta) = A_{1,n} + A_{2,n}$$

où $A_{1,n} = \boldsymbol{\delta}^t U_n$, $A_{2,n} = V_n$ telle que

$$U_n = \frac{-1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \psi_\tau(\epsilon_{i,\tau}), \quad V_n = \sum_{i=1}^n R\left(\epsilon_{i,\tau}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i\right).$$

Premièrement, vérifions que :

$$\mathbb{E}(\mathbf{x}_i \psi_\tau(\epsilon_{i,\tau})) = 0, \quad \text{Var}(\mathbf{x}_i \psi_\tau(\epsilon_{i,\tau})) = \tau(1 - \tau) \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t).$$

En effet :

Pour la moyenne, on a $\mathbb{E}(\mathbf{x}_i \psi_\tau(\epsilon_{i,\tau})) = \mathbb{E}(\mathbb{E}(\psi_\tau(\epsilon_{i,\tau}) | \mathbf{x}_i)) = 0$, puisque d'après l'hypothèse $q_\tau(\epsilon_{i,\tau} | \mathbf{x}_i) = 0$ et donc $\mathbb{E}(\rho'_\tau(\epsilon_{i,\tau} | \mathbf{x}_i)) = \mathbb{E}(\psi_\tau(\epsilon_{i,\tau} | \mathbf{x}_i)) = 0$.

Pour la variance, et puisque la moyenne est égale à zéro il suffit de calculer $\mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t \psi_\tau^2(\epsilon_{i,\tau}))$.

Notons d'abord que puisque $\mathbb{E}(\mathbb{I}(\epsilon_{i,\tau} < 0)) = \tau$, nous avons

$$\psi_\tau^2(\epsilon_{i,\tau}) = \tau^2 \mathbb{I}_{[0, +\infty[}(\epsilon_{i,\tau}) + (1 - \tau)^2 \mathbb{I}_{[-\infty, 0[}(\epsilon_{i,\tau})$$

Donc,

$$\begin{aligned} \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t \psi_\tau^2(\epsilon_{i,\tau})) &= \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t \mathbb{E}(\psi_\tau^2(\epsilon_{i,\tau}) | \mathbf{x}_i)) \\ &= \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t (\tau^2 \mathbf{P}(\epsilon_{i,\tau} > 0 | \mathbf{x}_i) + (1 - \tau)^2 \mathbf{P}(\epsilon_{i,\tau} \leq 0 | \mathbf{x}_i))) \\ &= \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t (\tau^2 (1 - \tau) + (1 - \tau)^2 \tau)) , \quad \text{puisque } \tau = F(\epsilon_{i,\tau} | \mathbf{x}_i) \\ &= \tau(1 - \tau) \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t) \end{aligned}$$

Maintenant, d'après le théorème central limite, on a :

$$U_n = \frac{-1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \psi_\tau(\epsilon) \xrightarrow{d} N(0, \tau(1 - \tau) \boldsymbol{\Sigma}).$$

Deuxièmement, pour le second terme $A_{2,n}$, on a :

$$\begin{aligned}
\mathbb{E} \left(R \left(\epsilon_{i,\tau}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i \right) \right) &= \mathbb{E} \left[\int_0^{\frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i} (\mathbb{1}(\epsilon_{i,\tau} \leq s) - \mathbb{1}(\epsilon_{i,\tau} \leq 0)) ds \right] \\
&= \mathbb{E} \left[\mathbb{E} \left[\int_0^{\frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i} (\mathbb{1}(\epsilon_{i,\tau} \leq s) - \mathbb{1}(\epsilon_{i,\tau} \leq 0)) ds \middle| \mathbf{x}_i \right] \right] \\
&= \mathbb{E} \left[\int_0^{\frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i} (F_{\epsilon_\tau|\mathbb{X}}(s) - F_{\epsilon_\tau|\mathbb{X}}(0)) ds \right].
\end{aligned}$$

En utilisant la formule Taylor de premier ordre pour $F_{\epsilon_\tau|\mathbb{X}}(s)$ au voisinage de 0, on trouve

$$\begin{aligned}
\mathbb{E} \left(R \left(\epsilon_{i,\tau}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i \right) \right) &\simeq \mathbb{E} \left[\int_0^{\frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i} f_{\epsilon_\tau|\mathbb{X}}(0) s ds \right] \\
&= f_{\epsilon_\tau|\mathbb{X}}(0) \frac{S^2}{2} \bigg|_0^{\frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i} \\
&= \frac{1}{2} f_{\epsilon_\tau|\mathbb{X}}(0) \boldsymbol{\delta}^t \left(\frac{1}{n} \mathbf{x}_i \mathbf{x}_i^t \right) \boldsymbol{\delta}.
\end{aligned}$$

Avec un développement similaire, on peut montrer que $\mathbb{V}ar(R(\epsilon_{i,\tau} - \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i)) \rightarrow 0$, quand n tend vers l'infini et donc, par l'inégalité de Tchybechev,

$$R \left(\epsilon_{i,\tau}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i \right) = \mathbb{E} \left(R \left(\epsilon_{i,\tau}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i \right) \right) + o_p(1),$$

c-à-d, pour tout ϵ , on

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| R \left(\epsilon_{i,\tau}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i \right) - \mathbb{E} \left(R \left(\epsilon_{i,\tau}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i \right) \right) \right| > \epsilon \right) = 0.$$

Donc,

$$\begin{aligned}
A_{2,n} &= \frac{1}{2} f_{\epsilon_\tau|\mathbb{X}}(0) \boldsymbol{\delta}^t \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \right) \boldsymbol{\delta} + o_p(1) \\
&= \frac{1}{2} f_{\epsilon_\tau|\mathbb{X}}(0) \boldsymbol{\delta}^t \boldsymbol{\Sigma} \boldsymbol{\delta} + o_p(1).
\end{aligned}$$

Par conséquent,

$$A_n(\boldsymbol{\delta}) = \boldsymbol{\delta}^t U_n + \frac{1}{2} f_{\epsilon_\tau|\mathbb{X}}(0) \boldsymbol{\delta}^t \boldsymbol{\Sigma} \boldsymbol{\delta} + o_p(1).$$

Par définition de $\hat{\beta}_\tau$ on a :

$$\begin{aligned}\hat{\beta} &= \operatorname{argmin}_{\beta_\tau} \sum_{i=1}^n \rho_\tau(y_i - \beta_\tau^t x_i) \\ &= \operatorname{argmin}_{\delta} \hat{A}_n \left(\beta_\tau + \frac{1}{\sqrt{n}} \delta \right).\end{aligned}$$

Donc, $\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) = \operatorname{argmin}_{\delta} \hat{A}_n(\delta)$. D'après le corollaire 1 de Hjort et Pollard [HP93], $\sqrt{n}(\hat{\beta} - \beta)$ et $-f_{\epsilon_\tau|\mathbb{X}}^{-1}(0)\Sigma^{-1}U_n$ suivent la même loi. Par conséquent,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N \left(0, f_{\epsilon_\tau|\mathbb{X}}^{-2}(0|\mathbb{X})\tau(1-\tau)\Sigma^{-1} \right).$$

■

Voici quelques avantages et points faibles de la régression quantile linéaire.

Avantages

- Ce modèle généralise l'idée de la régression médiane. Ceci permet d'avoir une description plus précise de l'effet des facteurs sur la distribution conditionnelle d'une variable d'intérêt comparée à une régression linéaire médiane. En fait, la distribution conditionnelle est caractérisée par $q_\tau, \forall \tau \in (0, 1)$.
- La robustesse des estimateurs de la régression quantile qui est un des points forts de cette régression. En effet, la régression quantile est plus basée sur les rangs des observations que sur leurs valeurs. Ce qui lui permet d'être moins sensible aux valeurs aberrantes.
- La régression linéaire quantile est facile à interpréter grâce à son lien avec la distribution conditionnelle.

Inconvénients

- L'estimation de la régression linéaire quantile ne possède pas de forme explicite à cause de la norme L_1 . L'estimation se fait par des algorithmes d'optimisation.

Avec le développement des logiciels (R, Python, SAS, etc), ce point n'est plus un grand inconvénient.

- Moins d'efficacité (de variabilité) de l'estimateur de la régression linéaire quantile en comparaison avec la MCO. La régression linéaire quantile exige assez de données pour avoir une estimation de bonne qualité.
- La régression linéaire quantile suppose une forme de dépendance linéaire à tout les niveaux τ . Ceci peut être corrigé en considérant une forme non-linéaire ou un modèle non-paramétrique qui suppose une régression linéaire quantile localement. Cette dernière est appelée la régression linéaire locale, voir [YJ98] et [FG96] pour plus de détails.

1.5.2 Régression expectile linéaire

Dans la section précédente, on a présenté et étudié la régression quantile qui est une généralisation de la régression médiane. Inspirée par cette idée, la régression linéaire expectile (RE) est une extension naturelle de la régression linéaire moyenne.

En présence d'une ou plusieurs variables explicatives, pour $\omega \in (0, 1)$ fixe, la régression expectile est définie par :

$$\mathbf{Y} = \mathbf{X}^t \boldsymbol{\beta}_\omega + \boldsymbol{\epsilon}_\omega \quad \text{avec} \quad \mu_\omega(\boldsymbol{\epsilon}_\omega | \mathbb{X}) = 0, \quad (1.10)$$

où $\boldsymbol{\beta}_\omega = (\beta_0^\omega, \beta_1^\omega, \dots, \beta_p^\omega)$ est le vecteur de paramètres du modèle correspondant à ω ,

$$\mu_\omega(Y | \mathbb{X}) = \underset{\boldsymbol{\beta}_\omega \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \mathbb{E}[r_\omega(Y - \mathbb{X}\boldsymbol{\beta}_\omega) | \mathbb{X}] \quad (1.11)$$

et la fonction de perte r_ω est définie par (1.6).

L'hypothèse $\mu_\omega(\boldsymbol{\epsilon}_\omega | \mathbb{X}) = 0$ garantit que l'erreur du modèle est centrée au niveau de son expectile de niveau ω . La régression expectile pour toutes les valeurs ω caractérise la distribution conditionnelle de Y étant donné \mathbb{X} .

Remarque 1.16 *Si l'hypothèse d'homoscédasticité est satisfaite, alors*

$$\mu_\omega(Y|\mathbb{X}) = (\beta_0^\omega + \mu_\omega(\epsilon_\omega|\mathbb{X})) + \beta_1^\omega x_{i,1} + \cdots + \beta_p^\omega x_{i,p}.$$

Par conséquent, seulement l'ordonnée à l'origine qui change avec ω . Et si $\mu_\omega(\epsilon_\omega|\mathbb{X}) = 0$, on aura des droites parallèles pour les différentes valeurs de ω .

Si on suppose un modèle hétéroscédastique, par exemple :

$$y_i = \beta_0^\omega + \beta_1^\omega x_{i,1} + \cdots + \beta_p^\omega x_{i,p} + (1 + \alpha x_{i,1})\epsilon_\omega,$$

alors

$$\mu_\omega(Y|\mathbb{X}) = (\beta_0^\omega + \mu_\omega(\epsilon_\omega|\mathbb{X})) + (\beta_1^\omega + \alpha\mu_\omega(\epsilon_\omega|\mathbb{X}))x_{i,1} + \cdots + \beta_p^\omega x_{i,p}.$$

et dans ce cas, l'ordonnée à l'origine et la pente correspondante à la variable $x_{i,1}$ changent avec ω . La figure illustre le cas de la régression linéaire expectile dans le cas hétéroscédastique.

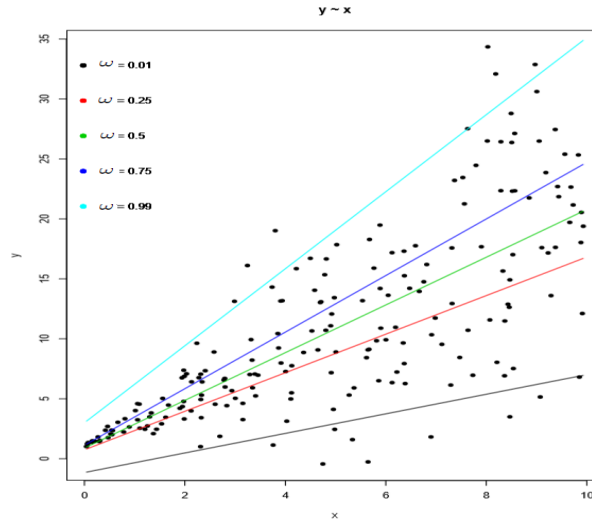


Figure 1.11 – Illustration de la régression expectile pour différents ω avec le modèle $Y = \beta_0^\omega + \beta_1^\omega x_{i,1} + \alpha x_{i,1}\epsilon_i$.

L'estimation de ce modèle est proposée par Newey et Powell [NP87] et consiste à minimiser les erreurs en utilisant la perte r_ω :

$$\hat{\beta}_\omega \equiv (\beta_0^\omega, \beta_1^\omega, \dots, \beta_p^\omega) = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^n r_\omega(y_i - \beta_0^\omega - \beta_1^\omega x_{i,1} - \dots - \beta_p^\omega x_{i,p}).$$

où $r_\omega(u) = u^2|\omega - \mathbb{1}(u \leq 0)| = u^2|\psi_\omega(u)|$ fonction de perte associée à la RE. Notons que $r'_\omega(u) = 2u|\psi_\omega(u)|$.

Contrairement à la fonction de perte L_1 , la fonction de perte L_2 est continument différentiable. Donc, en utilisant les conditions d'optimalité du premier ordre (KKT1), on peut montrer que l'estimateur est donné par :

$$\hat{\beta}_\omega = \left(\sum_{i=1}^n |\psi_\omega(\hat{\epsilon}_{i,\omega})| \mathbf{x}_i \mathbf{x}_i^t \right)^{-1} \sum_{i=1}^n |\psi_\omega(\hat{\epsilon}_{i,\omega})| y_i \mathbf{x}_i. \quad (1.12)$$

où $\hat{\epsilon}_{i,\omega} = y_i - \hat{\beta}_0^\omega - \hat{\beta}_1^\omega x_{i,1} - \dots - \hat{\beta}_p^\omega x_{i,p}$.

L'écriture matricielle de l'estimateur $\hat{\beta}_\omega$ est comme suit :

$$\hat{\beta}_\omega = (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} \mathbf{Y},$$

où $\mathbf{W} = \operatorname{diag}(|\psi_\omega(\hat{\epsilon}_{1,\omega})|, \dots, |\psi_\omega(\hat{\epsilon}_{n,\omega})|)$.

La formule (1.12) ne permet pas de trouver directement l'estimateur $\hat{\beta}_\omega$ puisque $\psi_\omega(\hat{\epsilon}_i)$ dépend de β_ω . Cependant, avec l'équation (1.12) et en utilisant un algorithme itératif, on peut évaluer $\hat{\beta}_\omega$. Par exemple, commencer l'algorithme avec l'estimateur par la méthode des MCO, β^{MCO} , pour calculer

$$\psi_\omega(\hat{\epsilon}_{i,\omega}) = \left| (\omega - \mathbb{1}(y_i - \hat{\beta}_0^{MCO} - \beta_1^{MCO} x_{i,1} - \dots - \beta_p^{MCO} x_{i,p})) < 0 \right|.$$

Ensuite, utiliser la formule (1.12) pour calculer $\hat{\beta}_\omega$. Recommencer l'itération et calculer $\hat{\epsilon}_{i,\omega}$ avec le nouveau estimateur de β_ω . Utiliser un critère d'arrêt pour obtenir l'estimateur des paramètres de la régression linéaire expectile.

Remarque 1.17 Comme pour la régression moyenne et quantile, l'estimation des paramètres du modèle de la régression linéaire expectile est équivalent à la maximisation de la vraisemblance en supposant la loi normale asymétrique, de paramètre (μ, σ^2, ω) , à la variable de réponse donnée par :

$$f(y|\mu, \sigma, \omega) = \frac{1}{\sqrt{2\pi}\sigma^2} \frac{\sqrt{2\omega(1-\omega)}}{\sqrt{\omega} + \sqrt{1-\omega}} \exp\left(-\rho_\omega\left(\frac{y-\mu}{\sigma}\right)\right).$$

Donc, en posant $\mu_{i,\omega} = \mathbf{x}_i^t \boldsymbol{\beta}_\omega$, la vraisemblance est de la forme suivante

$$\mathbb{L}(\boldsymbol{\beta}, \sigma | \mathbf{y}, \omega) \propto \sigma^{-n} \prod_{i=1}^n \exp\left(-\rho_\omega\left(\frac{y_i - \mathbf{x}_i^t \boldsymbol{\beta}_\omega}{\sigma}\right)\right).$$

Cette densité est différente de celle proposée par Azzalini1985 [Azz85]). Aussi, elle est différente de la densité connue sous le nom de AEPD (asymmetric exponential power distribution). Pour plus de propriétés sur cette dernière densité, voir [ZWZ09].

Newey et Powell [NP87] ont montré, sous certaines conditions de régularité, la normalité de $\widehat{\boldsymbol{\beta}}_\omega$.

Théorème 1.18 (Normalité asymptotique de $\widehat{\boldsymbol{\beta}}_\omega$) Sous les conditions 1-4 dans [NP87], nous avons :

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_\omega) \sim N(0, \mathbf{V}^{-1} \mathbf{D} \mathbf{V}^{-1}),$$

où

$$\mathbf{V} = \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t \psi_\omega(\epsilon_{i,\omega})), \quad \mathbf{D} = \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t \epsilon_{i,\omega}^2 \psi_\omega(\epsilon_{i,\omega})).$$

Démonstration

Ici, on donne une preuve différente de celle dans [NP87]. Notre preuve est inspirée par la preuve de la normalité de la régression linéaire quantile dans le théorème 1.15. Pour cela, On définit

$$H_n(\boldsymbol{\delta}) = \sum_{i=1}^n \left[r_\omega\left(\epsilon_{i,\omega} - \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i\right) - r_\omega(\epsilon_{i,\omega}) \right],$$

On peut montrer que :

$$r_\omega(u - v) - r_\omega(u) = -2uv\psi_\omega(u) + (2\omega - 1)R(u, v)(u - v) + v^2\psi_\omega(u), \quad (1.13)$$

où $R(u, v) = [\mathbb{1}(u \leq 0) - \mathbb{1}(u \leq v)](u - v)$ comme défini dans la preuve du théorème 1.15. En effet,

$$\begin{aligned} r_\omega(u - v) - r_\omega(u) &= |\omega - \mathbb{1}(u \leq v)|(u - v)^2 - |\omega - \mathbb{1}(u \leq 0)|u^2 \\ &= (|\omega - \mathbb{1}(u \leq v)| - |\omega - \mathbb{1}(u \leq 0)|)(u - v)^2 \\ &\quad + v^2|\omega - \mathbb{1}(u \leq 0)| - 2uv|\omega - \mathbb{1}(u \leq 0)| \\ &= -2uv\psi_\omega(u) + \underbrace{(|\omega - \mathbb{1}(u \leq v)| - |\omega - \mathbb{1}(u \leq 0)|)}_{(*)}(u - v)^2 + v^2\psi_\omega(u) \\ &= -2uv\psi_\omega(u) + (2\omega - 1)R(u, v)(u - v) + v^2\psi_\omega(u). \end{aligned}$$

puisque

$$(*) = \begin{cases} (1 - \omega) - \omega & u \leq v, u > 0 \\ (1 - \omega) - (1 - \omega) & u \leq v, u \leq 0 \\ \omega - (1 - \omega) & u \geq v, u \leq 0 \\ \omega - \omega & u \geq v, u > 0 \end{cases} = \begin{cases} -(2\omega - 1) & u \leq v, u > 0 \\ (2\omega - 1) & u \geq v, u \leq 0 \end{cases}$$

En posant $u = \epsilon_{i,\omega}$ et $v = \frac{1}{\sqrt{n}}\boldsymbol{\delta}^t \mathbf{x}_i$, la décomposition (1.13) nous permettra d'écrire $H_n(\boldsymbol{\delta})$ comme suit :

$$\begin{aligned} H_n(\boldsymbol{\delta}) &= \boldsymbol{\delta}^t U_n + \frac{1}{2}\boldsymbol{\delta}^t V_n \boldsymbol{\delta} + (2\tau - 1) \sum_{i=1}^n R\left(\epsilon_{i,\omega}, \frac{1}{\sqrt{n}}\boldsymbol{\delta}^t \mathbf{x}_i\right) \left(\epsilon_{i,\omega} - \frac{1}{\sqrt{n}}\boldsymbol{\delta}^t \mathbf{x}_i\right) \\ &\equiv H_{1,n} + H_{2,n} + H_{3,n} \end{aligned}$$

où

$$U_n = \frac{-2}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \epsilon_{i,\omega} \psi_\omega(\epsilon_{i,\omega}), \quad V_n = \frac{2}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \psi_\omega(\epsilon_{i,\omega}).$$

Commençons par montrer la normalité asymptotique du terme U_n . Puisque $\mu_\omega(\epsilon_{i,\omega}) = 0$ implique que $\mathbb{E}(r'_\omega(\epsilon_{i,\omega}|\mathbf{x}_i)) = 2\mathbb{E}(\epsilon_{i,\omega}|\psi_\omega(\epsilon_{i,\omega})|\mathbf{x}_i) = 0$. Et donc,

$$\text{Var}(\mathbf{x}_i \epsilon_{i,\omega} \psi_\omega(\epsilon_{i,\omega})) = \mathbb{E}(\mathbf{x}_i \mathbf{x}_i^t \epsilon_{i,\omega}^2 \psi_\omega^2(\epsilon_{i,\omega})).$$

Par conséquent, d'après le théorème central limite, on a :

$$U_n \xrightarrow{d} N(0, 4\mathbf{D}). \quad (1.14)$$

En utilisant la loi faible des grands nombres sur V_n , on trouve :

$$V_n = \frac{2}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \psi_\omega(\epsilon_{i,\omega}) \xrightarrow{P} 2\mathbf{V}. \quad (1.15)$$

Montrons maintenant que le dernier terme $H_{3,n}$ est négligeable. À partir de l'inégalité suivante : $|R(u, v)(u - v)| = R(u, v)|u - v| \leq R(u, v)|v|$ (voir [Koe05]).

$$\begin{aligned} (2\omega - 1) \sum_{i=1}^n \left| R(\epsilon_{i,\omega}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i) (\epsilon_{i,\omega} - \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i) \right| &\leq \frac{1}{\sqrt{n}} \overbrace{\sum_{i=1}^n R(\epsilon_{i,\omega}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i) |\boldsymbol{\delta}^t \mathbf{x}_i|}^{M} \\ &\quad \underbrace{O_p(1)} \\ &\leq \frac{M}{\sqrt{n}}. \end{aligned} \quad (1.16)$$

Le fait que $\sum_{i=1}^n R(\epsilon_{i,\omega}, \frac{1}{\sqrt{n}} \boldsymbol{\delta}^t \mathbf{x}_i) |\boldsymbol{\delta}^t \mathbf{x}_i| = o_p(1)$ dans le côté droit de la première inégalité est d'après [Koe05]. Mais, on peut le montrer en utilisant des étapes similaires pour le calcul de $A_{2,n}$ dans la preuve du théorème 1.15.

Alors, d'après (1.14), (1.15) et (1.16), on a

$$H_n(\boldsymbol{\delta}) = \boldsymbol{\delta}^t U_n + \frac{1}{2} \boldsymbol{\delta}^t V_n \boldsymbol{\delta} + o_p(1).$$

Par définition de $\hat{\boldsymbol{\beta}}_\omega$ on a :

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_\omega) = \hat{\boldsymbol{\delta}}^t = \underset{\boldsymbol{\delta}}{\operatorname{argmin}} H_n(\boldsymbol{\delta}).$$

En effet :

$$\begin{aligned} \epsilon_{i,\omega} - \frac{1}{\sqrt{n}} \hat{\boldsymbol{\delta}}^t \mathbf{x}_i &= y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}}_\omega \\ y_i - \mathbf{x}_i^t \boldsymbol{\beta}_\omega - \frac{1}{\sqrt{n}} \hat{\boldsymbol{\delta}}^t \mathbf{x}_i &= y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}}_\omega \\ \mathbf{x}_i (\hat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_\omega)^t &= \frac{1}{\sqrt{n}} \hat{\boldsymbol{\delta}}^t \mathbf{x}_i \\ \sqrt{n}(\hat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_\omega) &= \hat{\boldsymbol{\delta}}. \end{aligned}$$

Donc $\sqrt{n}(\hat{\beta}_\omega - \beta_\omega)$ minimise $\hat{H}_n(\delta)$. D'après le corollaire 1 de [HP93], $\sqrt{n}(\hat{\beta}_\omega - \beta_\omega)$ suit la même loi de $-\frac{U_n}{2V}$. Donc,

$$\sqrt{n}(\hat{\beta}_\omega - \beta_\omega) \xrightarrow{d} N(0, (2V)^{-1}4D(2V)^{-1}).$$

■

Les avantages et les inconvénients de la régression linéaire expectile ressemblent à ceux de la régression moyenne.

Avantages

- Le modèle de régression linéaire expectile est une généralisation de la régression de la moyenne ;
- Contrairement à la régression moyenne, la régression expectile caractérise la distribution conditionnelle de la variable d'intérêt ;
- Comme la méthode des MCO, l'estimation de la régression linéaire expectile est stable et donc efficace, c-à-d l'estimateur satisfait à l'égalité dans inégalité de Cramér-Rao, autrement dit, sa variance est égale à l'inverse de l'information de Fisher.
- Le calcul n'est plus un problème pour la régression quantile en raison des techniques d'optimisation avancées récentes, mais il est toujours vrai que le calcul dans la régression expectile est plus simple. Ainsi, le problème d'optimisation de la régression expectile peut être facilement résolu par de nombreux algorithmes de type IRLS (*iteratively reweighted least squares*).
- Une autre caractéristique intéressante de la régression expectile est que l'estimateur par régression par espérance dépend de la forme de la distribution entière, tandis que l'estimateur par régression quantile ne se fonde que sur les percentiles de la distribution de queue estimée. Par conséquent, l'estimateur de régression des

espérances contient des informations supplémentaires sur l’ampleur de la distribution de la queue et reflète la valeur réelle de façon plus précise, en particulier pour les distributions à queue lourde.

- La régression quantile est basée sur les rangs alors que la régression expectile est basée à la fois sur les rangs et les valeurs des observations

Inconvénients

- Contrairement à la régression quantile, la régression expectile est difficile à interpréter.
- En termes de robustesse, la régression quantile est plus résistante aux valeurs aberrantes que la régression expectile, car la régression quantile utilise la norme L_1 . Cependant, la sensibilité aux valeurs extrêmes peut être bénéfique si la détection de l’hétéroscédasticité dans les données est le principal intérêt.

1.6 Méthode lasso pour la régression expectile

Dans cette section on s’intéresse à l’approche de régularisation de la régression expectile qui nous permettra d’étudier la distribution conditionnelle sur toutes les expectiles de la variable d’intérêt et de sélectionner les variables indépendantes les plus importantes. Cette section est inspirée d’un travail conjoint avec Mehdi Karoui durant son stage d’étude à l’Université de Sherbrooke.

L’approche Lasso pour de la régression expectile consiste à minimiser le critère, par rapport à β_ω , suivant :

$$\mathbb{F}_\lambda(\beta_\omega) = \|\mathbf{W}^{1/2}(\mathbf{Y} - \mathbf{X}\beta_\omega)\|_2^2 + \lambda\|\beta_\omega\|_1, \quad \lambda \geq 0,$$

Minimiser la fonction de perte $\mathbb{F}_\lambda(\beta_\omega)$ revient à minimiser $\|\mathbf{W}^{1/2}(\mathbf{Y} - \mathbf{X}\beta_\omega)\|_2^2$ sous la contrainte $\|\beta_\omega\|_1 < t$ pour un t choisi. Autrement dit

$$\hat{\beta}_\omega(\lambda) = \underset{\beta_\omega \in \mathbb{R}^p}{\operatorname{argmin}} \|\mathbf{W}^{1/2}(\mathbf{Y} - \mathbf{X}\beta)\|_2^2 + \lambda \|\beta_\omega\|_1$$

1.7 Condition d'optimalité

On cherche à minimiser la fonction objectif $F_\lambda(\beta_\omega)$, dont on notera la solution optimale $\beta_\lambda(\omega)$. Trouver cette solution optimale reviendra à trouver la différentielle qui s'annule en un point $\hat{\beta}$, ce qui va nous permettre de déduire que ce point est la solution optimale de notre fonction. Mais comme cette dernière est non-différentiable à cause de la norme L_1 , on introduit la notion de sous-différentielle dans le but d'obtenir une caractérisation de la solution optimale. Pour avoir cette solution, un algorithme dit LARS (homotopique) voir [FR13] a été proposé dans le cas de la moyenne. Dans le travail en collaboration avec Mehdi Karoui et soumis pour publication à la revue Camus, nous avons développé un algorithme afin d'avoir la solution de la régression expectile Lasso.

La méthode homotopique calcule, pour ω fixe, toutes les solutions en commençant par fixer une grande valeur λ_0 et poser $\beta_{\lambda_0}(\omega) = 0$ jusqu'à l'obtention de $\beta_\lambda(\omega)$ pour une valeur donnée de λ ou jusqu'à ce qu'on ait plus d'unicité de la solution.

Pour effectuer ces calculs, nous aurons besoin des conditions d'optimalité de Karush-Kuhn-Tucker (KKT1) qui vont nous permettre de suivre l'évolution de $\beta_\lambda(\omega)$ en fonction de λ .

$$l \in \{1, 2, \dots, n\} \left\{ \begin{array}{ll} (\mathbf{X}^T \mathbf{W}(\mathbf{X}\beta_\omega - \mathbf{Y}))_l = -\lambda \operatorname{signe}(\beta_{\omega,l}) & \beta_{\omega,l} \neq 0 \\ |(\mathbf{X}^T \mathbf{W}(\mathbf{X}\beta_\omega - \mathbf{Y}))_l| \leq \lambda, & \beta_{\omega,l} = 0. \end{array} \right.$$

CHAPITRE 2

Régression linéaire pour les données censurées

Dans le chapitre précédent, on a étudié le cas où les variables réponses sont complètement observées. Cependant, souvent la variable d'intérêt est incomplète due, par exemple, aux données manquantes ou à la censure. Ici, nous allons traiter l'estimation de la régression quand la variable réponse n'est pas toujours observée à cause de la censure à droite.

En analyse de survie, dans plusieurs domaines (médecine, sociologie, ingénierie), on s'intéresse au temps pour qu'un évènement survienne. Par exemple, en médecine, on cherche à modéliser le temps de survie chez des patients souffrants d'une maladie. En économie, la durée de chômage est un indicateur crucial de l'inégalité sociale et de la bonne santé de l'économie. Enfin, en ingénierie, la durée de vie des machines est étudiée pour vérifier leurs qualités ou prévoir le temps de son remplacement. Dans ces études, on analyse un échantillon d'observations incomplètes dues à la censure. Les différentes méthodes sta-

tistiques pour les données complètes dans ce cas ne sont plus applicables et donnent des résultats souvent biaisés. Des nouvelles approches en analyse de survie ont été proposées. Elles nous permettent d'étudier les données censurées, c-à-d des données dont on connaît la borne (inférieure ou/et supérieure) et non une valeur précise. En particulier, la régression linéaire avec la variable réponse censurée à droite sera étudiée tout au long de ce chapitre. On supposera que les variables explicatives sont complètement observées, mais les méthodologies peuvent être adaptées pour considérer le cas des variables explicatives censurées. Pour plus de détails sur l'analyse de survie, voir Klein et Moeschberger [KM03].

2.1 Distribution de la durée de survie

Dans cette section, nous allons rappeler certaines fonctions utiles pour modéliser le temps de survie ainsi que la définition de la censure à droite. Soit Y , une variable non-négative, absolument continue par rapport à la mesure de Lebesgue et représentant une durée de survie. Souvent, la loi de Y est représentée par sa densité ou sa distribution. Mais en analyse de survie, la loi de probabilité de Y est généralement analysée en utilisant la fonction de survie, fonction de hasard instantanée ou fonction de hasard cumulée :

1. Fonction de survie :

La fonction de survie est la probabilité de survivre jusqu'à l'instant y c-à-d

$$S(y) = P(Y > y) = 1 - F(y),$$

où $F(y) = P(Y \leq y)$ est la fonction de répartition de Y .

On peut voir que $S(y)$ est une fonction décroissante sur l'intervalle $[0, 1]$.

2. Fonction hasard instantanée (taux de hasard) :

La fonction hasard en y caractérise la probabilité de mourir dans un intervalle $[y, y + h]$, étant donné que l'individu a survécu au temps y . Cette fonction possède différentes formes qui nous permettent de résumer les données de survie .

$$\begin{aligned} h(y) &= \lim_{h \rightarrow 0} \frac{P(Y \leq y + h | Y \geq y)}{h} \\ &= \frac{1}{P(Y \geq y)} \lim_{h \rightarrow 0} \frac{P(Y \leq y + h)}{h} \\ &= \frac{f(y)}{S(y)} \\ &= -\frac{d}{dy} \log S(y) \end{aligned}$$

où f est la densité de la variable Y . La fonction de hasard n'est pas nécessairement monotone et souvent préférée par les analystes puisqu'elle donne le risque instantané de mourir, par exemple ou tomber en panne pour une machine.

3. Fonction de hasard cumulée :

La fonction de hasard cumulée est l'intégrale de la fonction hasard $h(y)$:

$$H(y) = \int_0^y h(u) du = -\log(S(y)).$$

Par conséquent, la fonction de survie peut s'écrire sous la forme suivante :

$$S(y) = e^{-H(y)}.$$

On peut déduire que :

$$f(y) = h(y)e^{-H(y)}.$$

La fonction $H(y)$ est croissante dans l'intervalle $[0, +\infty)$

2.1.1 Censure à droite

Différentes sortes de censure ont été étudiées dans la littérature. Dans ce travail on s'intéresse à la censure à droite. Dans ce cas, au lieu d'avoir la variable d'intérêt Y on

observe le vecteur aléatoire $(T, \delta) = (\min(Y, C), 1_{\{Y < C\}})$ où C est la variable de censure. Cette dernière peut être fixe ou aléatoire ce qui nous ramène différents scénarios de censure à droite.

Censure de type I : Souvent dans les essais cliniques, on cherche à étudier la durée de survie en observant n éléments (souris, machines, etc) durant une période fixe c . Donc, dans ce cas, la censure est une valeur non-aléatoire et on peut observer l'événement durant la période $[0, c]$ sinon l'observation est censurée par c .

Censure de type II : Dans ce cas, on commence notre étude avec n éléments et on arrête l'expérience quand on observe l'événement (décès, rémission, panne, etc) pour un nombre fixe r d'individus ($r < n$) et, donc, les durées de vie des $(n - r)$ éléments sont censurées par la valeur aléatoire de la durée de vie du dernier élément qui a subi l'événement.

La censure aléatoire : La censure aléatoire à droite est le cas le plus fréquent. Dans ce cas, la censure est engendrée différentes situations, par exemple :

- fin de l'étude ;
- l'individu déménage dans un autre endroit et quitte l'étude définitivement ;
- changement du traitement ou l'inefficacité du traitement à cause des effets secondaires ;
- décès du patient à cause d'un accident.

Dans ce modèle, on observe le couple (T_i, δ_i) , $\forall i = 1, \dots, n$, où $T_i = \min(Y_i, C_i)$ est la durée observée et $\delta_i = 1_{\{Y_i \leq C_i\}}$ est l'indicateur de la censure où les censures C_1, \dots, C_n sont des variables aléatoires non-négatives i.i.d. La figure 2.1 illustre le cas de la censure aléatoire à droite.

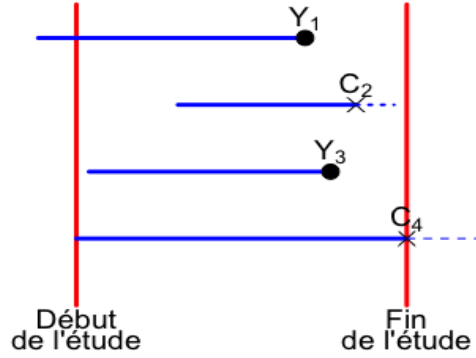


Figure 2.1 – Exemple de censure à droite aléatoire.

2.1.2 Estimateur de Kaplan-Meier

Dans cette section on cherche à estimer la fonction de survie dans le cas de la censure à droite. On considère un échantillon de taille n , (T_i, δ_i) , $\forall i = 1, \dots, n$, où $T_i = \min(Y_i, C_i)$ est la durée observée et $\delta_i = \mathbb{1}\{Y_i \leq C_i\}$. On supposera que C et Y sont indépendantes et on note respectivement par F et G la distribution de Y et de C et par S la fonction de survie de Y .

Dans le cas des données complètes, on estime souvent la fonction de survie par la fonction empirique, notée $S_n = 1 - F_n$ où F_n est la distribution empirique. Dans le cas de la censure à droite, l'estimateur S_n basé sur T_i ou sur $\{Y_i, \delta_i = 1\}$ n'est pas un estimateur convergent de S . En effet, si on considère S_n avec les T_i , on estimera la fonction suivante

$\mathbb{P}(T > y)$. Or,

$$\begin{aligned}
\mathbb{P}(T > y) &= \mathbb{P}(\min(Y, C) > y) \\
&= \mathbb{P}(Y > y, C > y) \\
&= \mathbb{P}(Y > y)\mathbb{P}(C > y) \text{ puisque les variables aléatoires } Y \text{ et } C \text{ sont indépendantes} \\
&= S(y)(1 - G(y)) \\
&\neq S(y).
\end{aligned}$$

De même, si on considère S_n avec les $\{\mathbf{Y}_i, \delta_i = 1\}$, on estimera $\mathbb{P}(T > y, \delta = 1)$. Or,

$$\mathbb{P}(T > y, \delta = 1) = \mathbb{P}(Y > y, Y \leq C) \leq \mathbb{P}(Y > y) = S(y).$$

Dans les deux cas, on sous-estimera S et ceci est naturel parce qu'on n'observe par les grandes valeurs de Y .

Dans le cas de la censure, Kaplan et Meier [KM58] ont proposé un estimateur appelé estimateur "Produit-Limite", pour la fonction de survie $S(t)$. C'est un estimateur efficace basé sur la méthode du maximum de vraisemblance.

Notations :

- d_i = nombre de décès au temps t_i , où t_i est une réalisation de T_i
- R_i nombre d'observations "à risque" (donc pas décédés ou censurés) juste avant l'instant t_i .

Notons que dans le cas d'absence d'ex-aequo,

$$d_i = \begin{cases} 0 & \text{pour } y_i \text{ censurée i.e quand } , \delta_i = 0 \\ 1 & \text{pour } y_i \text{ observée i.e quand } , \delta_i = 1 \end{cases}$$

Dans ce cas, l'estimateur de Kaplan-Meier, est donné par :

$$\begin{aligned}\hat{S}(t) &= \prod_{i=1, T_{(i)} \leq t}^n \left(1 - \frac{\delta_i}{R_i}\right) \\ &= \prod_{i=1, T_{(i)} \leq t}^n \left(1 - \frac{1}{n - i + 1}\right)^{\delta_{(i)}}\end{aligned}$$

où $T_{(1)} < T_{(2)} < \dots < T_{(n)}$ sont les statistiques d'ordre toutes différentes et $\delta_{(1)} < \delta_{(2)} < \dots < \delta_{(n)}$ sont les indicateurs correspondants.

La formule générale de l'estimateur Kaplan-Meier, si $d_i > 1$ et si on note $t_{(1)} < \dots < t_{(D)}$ les différents temps de la réalisation de l'événement est comme suit :

$$\hat{S}(t) = \begin{cases} \prod_{i=1, t_{(i)} \leq t}^n \left(1 - \frac{d_i}{R_i}\right) & \text{si } t \leq Y_{(n)} \\ 1 & \text{si } t > T_{(n)}. \end{cases}$$

La construction de cet estimateur en utilisant la méthode de la vraisemblance maximale est assez compliquée (voir [KM03]). Ici, on donnera une méthode alternative de la construction de l'estimateur de Kaplan-Meier. En effet, si $t'' < t' < t$, alors :

$$\begin{aligned}P(Y > t) &= P(Y > t', Y > t) \\ &= P(Y > t | Y > t') P(Y > t') \\ &= P(Y > t | Y > t') P(Y > t' | Y > t'') P(Y > t'')\end{aligned}$$

Alors, en considérant $T_{(0)} = 0$.

$$P(Y > T_{(j)}) = \prod_{k=0}^j P(Y > T_{(k)} | Y > T_{(k-1)}).$$

Or, la probabilité $p_i = P(Y \leq T_{(i)} | Y > T_{(i-1)})$ (probabilité de mourir dans l'intervalle $]T_{(i-1)}, T_{(i)}]$ sachant que le patient est en vie à l'instant $T_{(i-1)}$), peut-être estimée par $\hat{p}_i = \frac{d_i}{R_i}$.

Remarque 2.1 *Supposons que toutes les observations soient non censurées ($\delta_i = 1 \forall i$), l'estimateur de Kaplan-Meier est alors égal à la fonction de distribution empirique. En effet, considérons le cas d'absence d'ex-aequo, :*

$$\begin{aligned}
\hat{S}(t) &= \prod_{i=1, T_{(i)} \leq t, \delta_{(i)}=1}^n \left(1 - \frac{1}{n-i+1}\right) \\
&= \prod_{T_{(i)} \leq t} \frac{n-i}{n-i+1} \\
&= \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{n-j}{n-j+1} \quad \text{si } T_{(j)} \leq t < T_{(j+1)} \\
&= \frac{n-j}{n} \\
&= 1 - \frac{1}{n} \sum_{i=1}^n \mathbb{I}(T_i \leq t).
\end{aligned}$$

Exemple On génère $n = 200$ observations, Y_i , d'une densité exponentielle de paramètre $\lambda = 1$. Les observations censurées, C_i sont générées à l'aide d'une densité exponentielle de paramètre $p/(1-p)$ où p est le degré de censure. Dans cet exemple on considère $p = 0.5$ et on crée le couple $T_i = \min(Y_i, C_i)$ et $\delta_i = \mathbb{I}(Y_i \leq C_i)$. La figure 2.2 présente la vraie distribution, en bleue, l'estimateur de Kaplan-Meier, en noir et la distribution empirique avec les T_i en rouge. On voit clairement que l'estimateur KM est proche de la vraie distribution et que la distribution empirique basée sur les données observées sur-estime la vraie distribution.

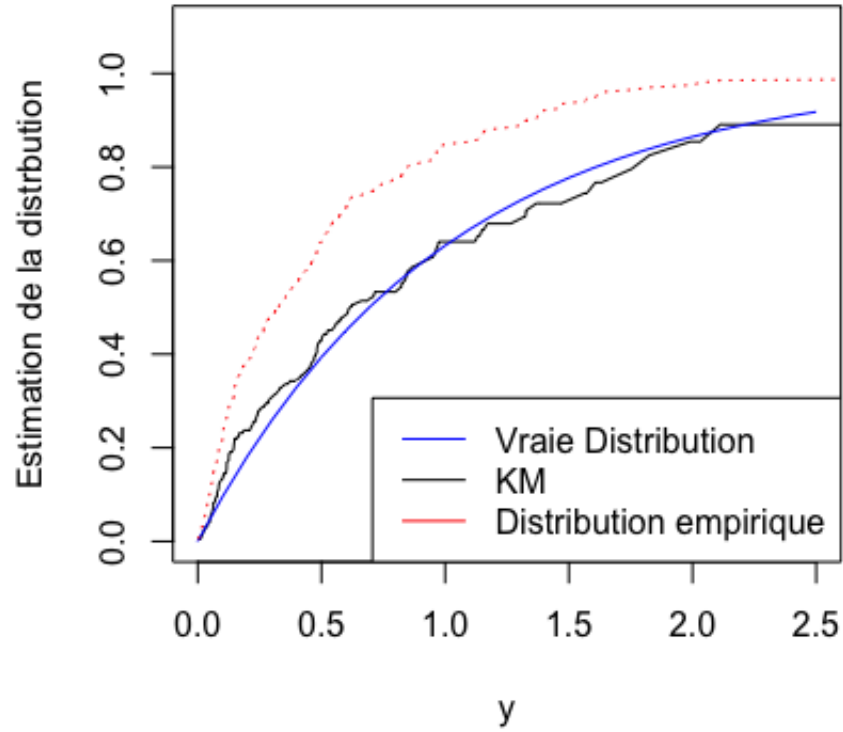


Figure 2.2 – Une illustration des estimateurs empiriques et Kaplan-Meier de la distribution.

2.1.3 Estimateur de Kaplan-Meier généralisé

En présence de covariables, la distribution de Y pourra changer selon les valeurs de ces covariables. Dans ce cas, on cherche à modéliser la distribution conditionnelle $F_{Y|\mathbb{X}}$ en utilisant un échantillon $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$. Pour les données complètes, Stone [Sto77]

a proposé en 1977 d'estimer $F_{Y|\mathbb{X}}$ par :

$$\widehat{F}_{Y|\mathbb{X}}(y) = \sum_{i=1}^n w_{ni}(\mathbf{x}, h) \mathbb{I}(y_i \leq y)$$

où $w_{n1}(x, h), \dots, w_{nn}(x, h)$ sont des fonctions de poids basées sur des noyaux appelés aussi kernel, et $h = h_n$ est le paramètre, appelé aussi fenêtre de lissage, qui dépend de la taille de l'échantillon. Le noyau permet d'accorder des poids aux observations (y_i, \mathbf{x}_i) , plus l'observation est proche de \mathbf{x} , en utilisant la distance L_2 par exemple, plus elle a un poids plus important et ce poids diminue plus on s'éloigne de cette valeur. Le paramètre contrôle la fenêtre des observations qui vont contribuer à l'estimation de la distribution. Donc, pour une petite valeur de h , on considère juste les observations très proches de \mathbf{x} et donc le biais est moins important, cependant, on risque d'avoir peu d'observations et donc une grande variance. Dans le cas contraire, pour une grande valeur de h , beaucoup de données vont être utilisées pour l'évaluation de la distribution donc une petite variance, cependant avec un grand biais. Le paramètre h permet de faire un compromis entre le biais et la variance. On peut déterminer le paramètre h à partir de la validation croisée. Pour plus de détails sur comment choisir ce paramètre, voir [HM85].

Le choix le plus utilisé des fonctions $w_{ni}(\mathbf{x}, h)$ est le poids de Nadaraya-Watson qui est donné par :

$$w_{ni}^{\text{NW}}(\mathbf{x}, h) = \frac{K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)}{S_{n,0}(\mathbf{x}, h)}. \quad (2.1)$$

où $K(u_1, \dots, u_p) = \prod_{j=1}^p k(u_j)$ est le produit des fonctions de densités symétriques et dérivables sur $[-1, 1]$ et $S_{n,0}(\mathbf{x}, h) = \sum_{i=1}^n K\{(\mathbf{x}_i - \mathbf{x})/h\}$. Plusieurs candidats pour k sont possibles. Par exemple les noyaux d'Epanechnikov et "Triweight", illustrés par la figure 2.3, sont donnés respectivement par :

$$K(y) = \frac{3}{4}(1 - y^2)^2 \mathbb{I}(|y| \leq 1)/32, \quad K(y) = \frac{35}{32}(1 - y^2)^3 \mathbb{I}(|y| \leq 1).$$

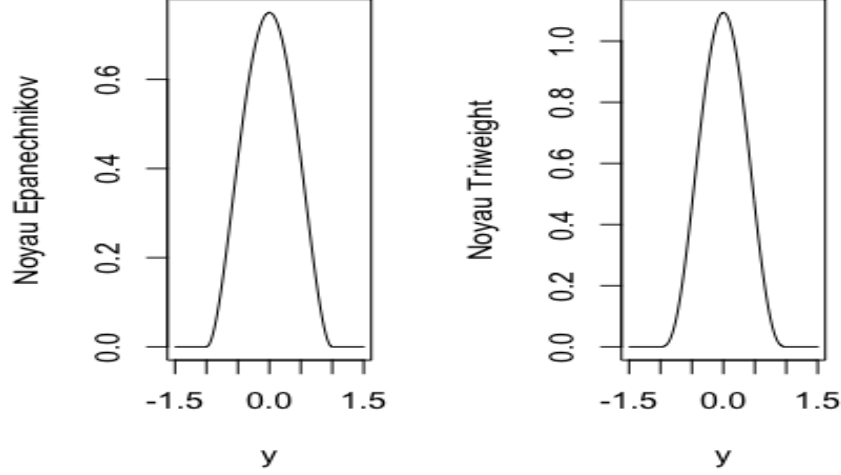


Figure 2.3 – Les noyaux d'Epanechnikov et "Triweight", .

En présence des données censurées et au point \mathbf{x}_i , on n'observe pas Y_i mais plutôt la variable $T_i = \min(Y_i, C_i)$ et l'indicatrice $\delta_i = \mathbb{1}\{Y_i \leq C_i\}$ où C_1, \dots, C_n sont les variables censurées indépendantes et non-négatives. On supposera, conditionnel à \mathbf{x}_i , Y_i et C_i sont indépendantes. Donc, la fonction de survie, $S(t|\mathbf{x})$, de Y étant donné $\mathbf{X} = \mathbf{x}$ satisfait la relation suivante :

$$(1 - F(t|\mathbf{x}_i)) = S(t|\mathbf{x}_i)(1 - G(t|\mathbf{x}_i))$$

où $F(.|\mathbf{x}_i)$ et $G(.|\mathbf{x}_i)$ sont, respectivement, les fonctions de distribution de T_i et C_i étant donné $\mathbf{X} = \mathbf{x}$.

L'estimateur de la fonction de survie $S(t|x)$ appelé "Estimateur de Kaplan-Meier généralisé (KMG)" est proposé par Beran [Ber81] et donné par :

$$\hat{S}_h(t|x) = \prod_{T_{(i)} \leq t} \left[1 - \frac{\omega_{n(i)}(\mathbf{x}, h)}{\sum_{i=1}^n \omega_{n(i)}(\mathbf{x}, h)} \right]^{\delta_{(i)}} \mathbb{I}(t < T_{(n)}). \quad (2.2)$$

Le temps d'évènement T_i est ordonné, $T_{(1)} < T_{(2)} < \dots < T_{(n)}$, les indicatrices $\delta_{(1)}, \dots, \delta_{(n)}$ et les poids $\omega_{n(i)}(\mathbf{x}, h), \dots, \omega_{n(n)}(\mathbf{x}, h)$ correspondent à δ_i et $\omega_{ni}(\mathbf{x}, h)$ respectivement.

Remarque 2.2 Si $\omega_{i,n}(\mathbf{x}, h) = \frac{1}{n}$, alors $\hat{S}_h(t|\mathbf{x})$ est l'estimateur classique de Kaplan – Meier étudié dans la partie précédente.

2.2 Estimation de la régression quantile et expectile : cas des données censurées

Dans cette section, nous allons étudier l'estimation de la fonction de régression quand la variable d'intérêt est censurée. Comme dans le chapitre 1, on considère le modèle linéaire $Y_i = \mathbf{x}_i^t \boldsymbol{\beta} + \epsilon_i$. Notons que dans le cas d'une durée de vie Z , la variable réponse est $Y = \log(Z)$. L'estimation de ses paramètres dans le cas des observations complètes sachant \mathbb{X} est donnée par :

$$m(x) = \underset{a}{\operatorname{argmin}} \mathbb{E}(\zeta(Y - a) | \mathbb{X}),$$

où ζ est une fonction de perte associée à une régression donnée. En supposant le modèle linéaire, la fonction $m(x)$ peut être estimée par :

$$\hat{m}(x) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^n \zeta(y_i - \mathbf{x}_i^t \boldsymbol{\beta}).$$

En présence des données censurées, cette dernière estimation de la régression n'est pas applicable, car nous n'observons pas toutes les valeurs de la variable réponse y_i , mais on observe le triplet $(T_i, \delta_i, \mathbf{x}_i)$, $\forall i = 1, \dots, n$, tel que $T = \min(Y, C)$, $\delta = \mathbb{1}(Y \leq C)$, où C est la variable de censure qui est, conditionnellement à \mathbb{X} , indépendante de la variable Y .

Afin de prendre en considération la censure C dans l'estimation des paramètres de la régression, on démontre résultat suivant :

Pour tout φ , fonction mesurable telle que $\varphi : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Alors, on a :

$$\mathbb{E}(\varphi(Y)|\mathbb{X}) = \mathbb{E} \left(\varphi(T) \frac{\delta}{1 - G_C(T^-|\mathbb{X})} \middle| \mathbb{X} \right) \quad (2.3)$$

où $G_C(c^-|\mathbf{x}) = \mathbb{P}(C < c|\mathbb{X})$ est la distribution de la variable de censure conditionnelle à \mathbb{X} .

Démonstration En effet,

$$\begin{aligned} \mathbb{E} \left(\varphi(T) \frac{\delta}{1 - G_C(T^-|\mathbb{X})} \middle| \mathbb{X} \right) &= \mathbb{E} \left(\varphi(Y) \frac{\delta}{1 - G_C(Y^-|\mathbb{X})} \middle| \mathbb{X} \right) \\ &= \mathbb{E} \left(\mathbb{E} \left(\varphi(Y) \frac{\delta}{1 - G_C(Y^-|\mathbb{X})} \middle| \mathbb{X}, Y \right) \right) \\ &= \mathbb{E} \left(\frac{\varphi(Y)}{1 - G_C(Y^-|\mathbb{X})} \mathbb{E}(\delta|\mathbb{X}, Y) \middle| \mathbb{X} \right). \end{aligned}$$

Or $\mathbb{E}(\delta|\mathbb{X}, Y) = \mathbb{P}(Y \leq C|\mathbb{X}, Y) = \mathbb{P}(Y \leq C|\mathbb{X})$ à cause de l'indépendance conditionnelle à \mathbb{X} de C et Y . Donc,

$$\begin{aligned} \mathbb{E} \left(\varphi(T) \frac{\delta}{1 - G_C(T^-|\mathbb{X})} \middle| \mathbb{X} \right) &= \mathbb{E} \left(\frac{\varphi(Y)}{1 - G_C(Y^-|\mathbb{X})} (1 - G_C(Y^-|\mathbb{X})) \middle| \mathbb{X} \right) \\ &= \mathbb{E}(\varphi(Y)|\mathbb{X}). \end{aligned}$$

Basées sur des choix de φ , des méthodes d'estimation de plusieurs fonctions sont proposées dans la littérature en présence des données censurées. Par exemple, si on considère $\varphi_t(z) = \mathbb{I}(z > t)$, on peut estimer la fonction de survie conditionnelle en utilisant

$$\mathbb{E} \left(\frac{\delta \mathbb{I}(T > t)}{1 - G_C(T^-|\mathbf{x})} \middle| \mathbf{X} = \mathbf{x} \right).$$

Par conséquent, un estimateur de la fonction de survie conditionnelle est donnée par :

$$\tilde{S}(t|\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{1 - \hat{G}_C(T^-|\mathbf{x})} \mathbb{I}(T_i > t)$$

où \widehat{G}_C est l'estimateur de Kaplan-Meier généralisé de la distribution de la censure.

Pour la variable réponse censurée, [Mil76] a proposé d'estimer la régression moyenne (voir aussi [BJ79]) et [ZYW95] ont étudié l'estimation de la régression médiane. Dans la section 2.2.1, nous allons présenter l'estimation de la régression quantile qui est très bien étudié dans la littérature. Pour une bonne revue dans la littérature sur la régression quantile quand la variable d'intérêt est censurée à droite, voir [FW06]. Dans la section 2.2.2, nous introduisons la régression expectile et son estimation. Cette méthode, à notre connaissance, n'est jamais été proposée ou étudiée pour les données censurées à droite.

2.2.1 Estimation de la régression quantile

On considère le modèle de régression linéaire quantile $Y_i = \mathbf{x}_i^t \boldsymbol{\beta} + \epsilon_i$ où le vecteur des covariables \mathbb{X} est observé et la variable dépendante Y peut-être censurée à droite. Pour $\tau \in (0, 1)$, on suppose que $\rho_\tau(\epsilon_i | \mathbb{X}) = 0$ avec $\rho(u) = u(\tau - \mathbb{1}(u < 0))$ est la fonction de perte associée à la régression quantile.

En supposant ce modèle et en utilisant l'équation (2.3), l'estimation des paramètres de la régression associée au quantile conditionnel $q_\tau(Y | \mathbb{X}) = \inf\{F_{Y|\mathbb{X}} \geq \tau\}$, est donnée par :

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta}_\tau \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^n \left[\rho_\tau(T_i - \mathbf{x}_i^t \boldsymbol{\beta}_\tau) \widehat{W}_i(\mathbf{x}_i) \right] \quad (2.4)$$

où $\widehat{W}_i(\mathbf{x}_i) = \frac{\delta_i}{1 - \widehat{G}_C(T_i^- | \mathbf{x}_i)}$ où l'estimateur, \widehat{G}_C , est l'estimateur de la fonction de distribution de la censure proposé par Beran en 1981 [Ber81] donné par :

$$\hat{G}_C(t | \mathbf{x}) = 1 - \prod_{c_{(i)} \leq t} \left[1 - \frac{\omega_{n(i)}(\mathbf{x}, h)}{\sum_{i=1}^n \omega_{n(i)}(\mathbf{x}, h)} \right]^{1 - \delta_{(i)}} I(t < c_{(n)}).$$

Dans cet estimateur de Kaplan-Meier généralisé, les censures C_i sont ordonnées, $C_{(1)} < C_{(2)} < \dots < C_{(n)}$ et où $\delta_{(1)}, \dots, \delta_{(n)}$ et $\omega_{n(i)}(x, h), \dots, \omega_{n(n)}(x, h)$ sont, respectivement, les

indicatrices et les poids correspondants à cet ordre. Dans cette dernière estimation, les observations censurées sont considérées comme observées et les valeurs réellement observées comme censurées, d'où la nécessité de remplacer $1 - \delta_i$ au lieu δ_i dans (2.2).

La solution de la régression quantile n'est pas explicite à cause de la valeur absolue (norme L_1). Mais plusieurs méthodes numériques ont été suggérées pour générer la solution telle que l'algorithme du simplexe. Aussi, on peut considérer le problème (2.4) comme étant la régression quantile pondérée. En pratique, le package *quantreg* implémenté sous le logiciel *R* par (Portnoy et Koenker 1997) peut générer une solution à ce problème, avec l'argument poids $\hat{W}_i(x) = \frac{\delta}{1 - \hat{G}_C(y_i^-|\mathbf{x})}$, telle que $\hat{G}_C(Y^-|\mathbf{x})$ est trouvée par l'estimateur de *KMG* de [Ber81].

Sous certaines conditions, [CT13] ont montré la convergence en loi (distribution) de l'estimateur $\hat{\beta}_\tau$. Soient $f_{Y|\mathbb{X}}$ et $g_{Y|\mathbb{X}}$ les densités, conditionnelles à \mathbb{X} , de Y et C , et $F_{Y|\mathbb{X}}$ et $G_{C|\mathbb{X}}$ les distributions de Y et C étant donné \mathbb{X} . Ainsi, lorsque n tend vers l'infini, on a :

$$\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) \sim N(0, \Gamma_1^{-1} \mathbf{V} \Gamma_1^{-1}),$$

où

$$\Gamma_1^{-1} = \mathbb{E} [\mathbf{X}^t \mathbf{X} f_{Y|\mathbf{X}=\mathbf{x}}(\mathbf{X} \beta_\tau)]$$

et où $\mathbf{V} = \text{Cov}(m_i(\beta_\tau, G) - (1 - \tau)\mathbf{X}_i f_{\mathbf{X}}(\mathbf{X}_i) \eta(T_i, \delta_i, \mathbf{X}_i^t \beta_\tau, \mathbf{X}_i))$

telle que $f_{\mathbf{X}}$ désigne la densité de \mathbf{X} , $m_i(\beta_\tau, G) = \mathbf{x}_i \left\{ \frac{\mathbb{1}(T_i \geq \mathbf{X} \beta_\tau)}{1 - G(\mathbf{X} \beta_\tau | \mathbf{x}_i)} - (1 - \tau) \right\}$

et $\eta(T_i, \delta_i, t, \mathbf{x}_i) = \int_0^{\min(T_i, t)} \frac{-g(s|\mathbf{x})ds}{\{1 - G(s|\mathbf{x})\}^2 \{1 - F(s|\mathbf{x})\}} + \frac{(1 - \delta_i) \mathbb{1}(T_i \leq t)}{1 - G(T_i|\mathbf{x}) \{1 - F(T_i|\mathbf{x})\}}.$

■

Pour plus de détails sur la démonstration voir [CT13].

Remarque 2.3 *Noter que dans le cas des données complètes, on a $\delta_i = 1$,*

$$\eta(T_i, \delta_i, t, \mathbf{x}_i) = 0.$$

Par conséquent,

$$\begin{aligned} \mathbf{V} &= \text{Cov}(m_i(\boldsymbol{\beta}_\tau, G)) \\ &= \mathbf{x}_i \{ \mathbb{1}(\epsilon_i \geq 0) - (1 - \tau) \} \quad \text{où} \quad \epsilon_i = y_i - \mathbf{X}_i \boldsymbol{\beta}_\tau \\ &= \mathbf{x}_i \psi_\tau(\epsilon_i). \end{aligned}$$

où $\psi_\tau(u) = \tau - \mathbb{1}(u \leq 0)$. Par conséquent, on retrouve le résultat énoncé dans le théorème 1.15 en remarquant que $f_{\epsilon|\mathbf{x}}(0) = f_{Y|\mathbf{X}}(\mathbf{X}\boldsymbol{\beta}_\tau)$.

2.2.2 Estimation de la régression expectile

En présence des données censurées, la méthode classique d'estimation de la RE vue au chapitre 2 n'est pas appropriée, à cause de la censure. La régression expectile quand la variable réponse est censurée n'a jamais été traitée à notre connaissance dans la littérature. Dans ce travail, nous proposons une nouvelle méthodologie pour estimer le modèle linéaire et ensuite présenter un algorithme pour évaluer cet estimateur.

Soit le modèle de régression linéaire expectile $y_i = \mathbf{x}_i^t \boldsymbol{\beta} + \epsilon_i$ et pour une valeur fixe $\omega \in (0, 1)$, on suppose que $r_\omega(\epsilon_i|\mathbb{X}) = 0$ où $r_\omega(u) = u^2|\tau - \mathbb{1}(u < 0)|$, tel que défini dans la section 1.5.2. En se basant sur la formule (2.3) avec $\varphi(T) = r_\omega(T - \mathbb{X}\boldsymbol{\beta}_\omega|\mathbb{X})$, On propose d'estimer les paramètres de ce modèle par :

$$\hat{\boldsymbol{\beta}}_\omega = \underset{\boldsymbol{\beta}_\omega \in \mathbb{R}^{p+1}}{\text{argmin}} \sum_{i=1}^n r_\omega(T_i - \mathbf{x}_i^t \boldsymbol{\beta}_\omega) \widehat{W}_i(\mathbf{x}_i). \quad (2.5)$$

où comme dans la section précédente, $\widehat{W}_i(\mathbf{x}) = \frac{\delta_i}{1 - \widehat{G}_C(y_i^-|\mathbf{x})}$ où l'estimateur, \widehat{G}_C est l'estimateur de la fonction de distribution de la censure proposé par [Ber81].

Le calcul différentiel (l'application des conditions du premier ordre) nous permettra d'obtenir l'estimateur du modèle de la RE. Rappelons que $r'_\omega(u) = -2u|\psi_\omega(u)|$ tel que $\psi_\omega(u) = (\omega - \mathbb{1}(u < 0))$. La solution du problème de minimisation est générée, en résolvant le système à $p + 1$ équations suivantes :

$$\begin{aligned} \frac{\partial L(\beta)}{\beta_\omega^0} \Big|_{\beta=\hat{\beta}_\omega} &= \sum_{i=1}^n \widehat{W}_i(\mathbf{x}) |\psi(\hat{\epsilon}_{i,\omega})| \hat{\epsilon}_{i,\omega} = 0 \\ \frac{\partial L(\beta)}{\beta_\omega^1} \Big|_{\beta=\hat{\beta}_\omega} &= \sum_{i=1}^n \widehat{W}_i(\mathbf{x}) x_{i,1} |\psi(\hat{\epsilon}_{i,\omega})| \hat{\epsilon}_{i,\omega} = 0 \\ &\vdots \\ \frac{\partial L(\beta)}{\beta_\omega^p} \Big|_{\beta=\hat{\beta}_\omega} &= \sum_{i=1}^n \widehat{W}_i(\mathbf{x}) x_{i,p} |\psi(\hat{\epsilon}_{i,\omega})| \hat{\epsilon}_{i,\omega} = 0 \end{aligned}$$

où $\hat{\epsilon}_{i,\omega} = T_i - (\hat{\beta}_\omega^0 + \hat{\beta}_\omega^1 x_{i,1} + \dots + \hat{\beta}_\omega^p x_{i,p})$. On pose la fonction de poids suivante :

$$\hat{\omega}_i(\mathbf{x}) = \widehat{W}_i(\mathbf{x}) |\psi_\omega(\hat{\epsilon}_{i,\omega})|.$$

En réorganisant les équations, on trouve :

$$\begin{aligned} \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) T_i &= \hat{\beta}_\omega^0 \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) + \hat{\beta}_\omega^1 \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,1} + \dots + \hat{\beta}_\omega^p \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,p} = 0 \\ \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,1} T_i &= \hat{\beta}_\omega^0 \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,1} + \hat{\beta}_\omega^1 \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,1} x_{i,1} + \dots + \hat{\beta}_\omega^p \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,1} x_{i,p} = 0 \\ &\vdots \\ \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,p} T_i &= \hat{\beta}_\omega^0 \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,p} + \hat{\beta}_\omega^1 \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,p} x_{i,1} + \dots + \hat{\beta}_\omega^p \sum_{i=1}^n \hat{\omega}_i(\mathbf{x}) x_{i,p} x_{i,p} = 0 \end{aligned}$$

Ce qui est équivalent à la forme matricielle suivante :

$$\mathbf{X}^t \mathbf{W} \mathbf{Y} = \mathbf{X}^t \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}_\omega,$$

où la matrice \mathbf{W} est une matrice diagonale s'écrit comme suit :

$$\mathbf{W} = \text{diag}(\hat{\omega}_1(\mathbf{x}), \hat{\omega}_2(\mathbf{x}), \dots, \hat{\omega}_n(\mathbf{x})).$$

Si on suppose que $\mathbf{X}^t \mathbf{W} \mathbf{X}$ est une matrice inversible, on déduit que :

$$\hat{\beta}_\omega = (\mathbf{X}^t \mathbf{W} \mathbf{Y})^{-1} \mathbf{X}^t \mathbf{W} \mathbf{X}.$$

Remarque 2.4 *La normalité asymptotique de cet estimateur est assez compliquée à cause du terme aléatoire $\widehat{W}_i(\mathbf{x})$. Par simulation dans la section 4, on voit clairement que la loi asymptotique de cet estimateur ressemble à une normale.*

Exemple Pour illustrer l'effet de la correction de l'estimation de la régression expectile et quantile pour $\tau = 0.5$, on génère $n = 100$ couples d'un modèle linéaire $Y_i = 2 + 4X_i + \epsilon_i$ où $X_i \sim \text{Uniforme}(0, 1)$ et $\epsilon_i \sim N(0, 1)$. Ensuite, on applique la censure en générant la variable de censure à partir d'une uniforme $[0, \lambda]$. La valeur de λ contrôle le degré de censure. Ici, on considère $\lambda = 16$ qui assure $p = 25\%$ de censure. La figure 2.4 illustre cet exemple où on a dessiné les vraies données avec la régression moyenne et la régression quantile avec la correction (les droites continues) et la méthode MCO sans correction (les droites discontinues). On voit clairement que sans la correction, la MCO sous estime la vraie régression due au fait que les grandes valeurs de Y sont en général censurée. Avec la correction, la régression moyenne et médiane donnent des résultats similaires.

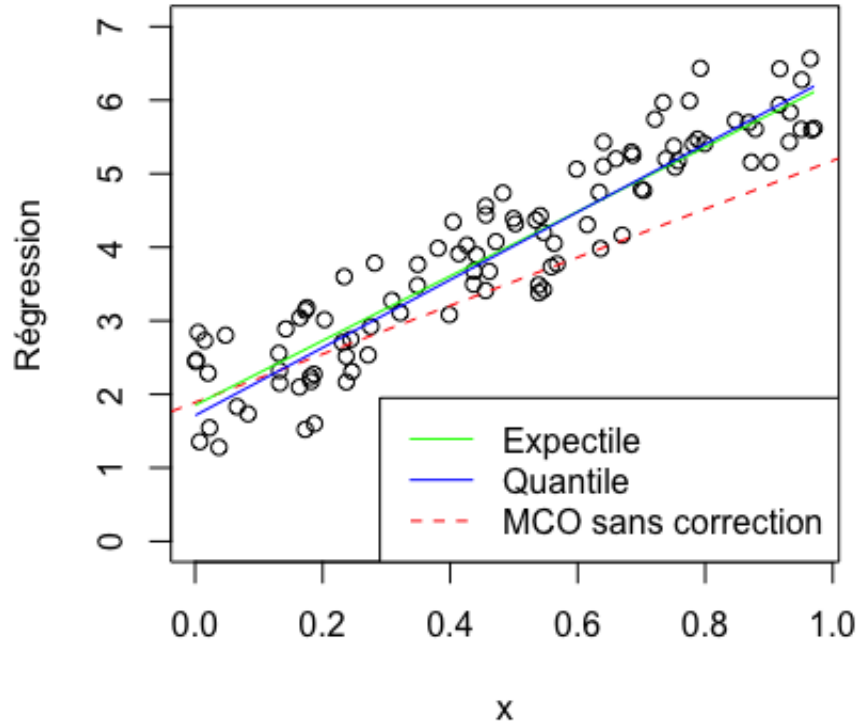


Figure 2.4 – Illustration d’une relation linéaire entre la variable réponse censurée et une variable explicative. Trois méthodes d’estimation : MCO et la régression quantile avec la correction (les droites continues) et la méthode MCO sans correction (les droites discontinues).

Voici un algorithme pour calculer l’estimateur de la régression expectile avec les données censurées.

Algorithme

Dans chaque itération, on calcule les valeurs $\widehat{W}_i(x) = \frac{\delta}{1 - \hat{G}_C(T_i^-|x)}$ à partir de KMG.

— **Itération 0**

- On calcule les estimateurs $\hat{\beta}_\omega^{(0)}$ par les méthodes des moindres carrés ordinaires MCO.
- On calcule ensuite les résidus $\hat{\epsilon}_{i,\omega}^{(0)}$, $i = 1, \dots, n$
- On calcule ensuite les poids $\hat{\omega}_i^{(0)} = \widehat{W}_i(\mathbf{x})\psi_\omega(\hat{\epsilon}_{i,\omega}^{(0)})$
- On obtient alors la matrice $\mathbf{W}^{(0)} = \text{diag}(\hat{\omega}_1^{(0)}, \hat{\omega}_2^{(0)}, \dots, \hat{\omega}_n^{(0)})$

— **Itération k**

- On calcule les estimateurs $\hat{\beta}_\omega^{(k)} = (\mathbf{X}^t \mathbf{W}^{(k-1)} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}^{(k-1)} \mathbf{Y}$.
- On calcule ensuite les résidus $\hat{\epsilon}_{i,\omega}^{(k)}$, $i = 1, \dots, n$
- On calcule ensuite les poids $\hat{\omega}_i^{(k)} = \widehat{W}_i(\mathbf{x})\psi_\omega(\hat{\epsilon}_{i,\omega}^{(k)})$
- On obtient alors la matrice $\mathbf{W}^{(k)} = \text{diag}(\hat{\omega}_1^{(k)}, \hat{\omega}_2^{(k)}, \dots, \hat{\omega}_n^{(k)})$

— **Arrêt de l'algorithme**

- On arrête le processus à la L^{me} itération lorsque l'algorithme converge :

$$\frac{\|\hat{\beta}_\omega^{(k)} - \hat{\beta}_\omega^{(k-1)}\|_2}{\|\hat{\beta}_\omega^{(k)}\|_2} < \xi \text{ où } \xi \text{ est un très petit nombre fixé.}$$

CHAPITRE 3

Approche robuste et efficace pour l'estimation de la régression linéaire

3.1 Introduction

La régression linéaire classique a pour but de modéliser et étudier la relation statistique entre la variable réponse et une ou plusieurs variables explicatives. Souvent, cette relation est résumée en étudiant l'effet des variables explicatives sur la moyenne de la distribution de la variable d'intérêt. Au cours de ces dernières décennies, cette modélisation a été développée en introduisant d'autres types de modèles de régression, telles que la régression quantile (RQ) introduite par Koenker et Bassett [KB78] et la régression expectile (RE) proposée par Newey et Powell [NP87]. Ces régressions estiment respectivement les quantiles et les expectiles de la fonction de répartition de la variable d'intérêt en fonction des variables explicatives (régresseurs). Chacun de ces deux modèles RE et RQ pos-

sède ses propres avantages et inconvénients au niveau de l'efficacité et de la robustesse. L'estimation de la régression expectile, qui généralise la régression moyenne, est stable, mais sensible aux points aberrants tandis que la méthode d'estimation de la régression quantile, qui généralise l'aspect de la médiane, est plus robuste par rapport aux points aberrants, mais pas efficace.

Dans ce chapitre nous allons introduire une nouvelle méthode d'estimation pratique qui permet d'avoir un bon compromis entre l'efficacité et la robustesse. Cette approche sera appelée, la régression HB (Hamdi et Bouezmarni). L'idée a été suggérée dans le travail d'Efron(1991) [Efr91], mais, à notre connaissance, jamais été étudiée.

Ce travail est réalisé en deux parties dans le cadre de la régression linéaire. La première est dans le cadre des données complètes et la deuxième en présence des données censurées. Dans la première partie, nous allons introduire l'estimateur HB et établir quelques propriétés asymptotiques. Dans la deuxième partie, nous allons introduire l'estimateur sans résultats asymptotiques vu la complexité de l'estimateur à cause de la censure.

3.2 Estimation HB de la régression linéaire pour les données complètes

Dans cette section, nous allons introduire l'estimation de la régression HB dans le cadre linéaire classique. On supposera que les données sont complètes.

Comme dans les autres chapitres, nous considérons un échantillon issu d'un modèle linéaire :

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta} + \epsilon_{i,\tau} \quad (3.1)$$

où y_i est la réponse de la $i^{\text{ème}}$ observation, $\mathbf{x}_i^t = (1, x_{i,1}, \dots, x_{i,p})$ est le vecteur composé de 1 et les p prédicteurs, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)$ est le vecteur des paramètres inconnus qu'on souhaite estimer et la variable $\epsilon_{i,\tau}$ est l'erreur du modèle.

Pour un $\tau \in (0, 1)$ et $1 \leq pk \leq 2$, l'estimation de la régression linéaire quantile et expectile est donnée par :

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^n \zeta_{\tau}^k(y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_p x_{i,p}). \quad (3.2)$$

où $\zeta_{\tau}^k(u) = |u|^k |\psi_{\tau}(u)|$ avec $\psi_{\tau}(u) = |\tau - \mathbb{1}_{(u \leq 0)}|$ avec $k = 1$ pour la régression quantile et $k = 2$ pour la régression expectile. D'après Efron (1989), plus p est proche de 1, plus l'estimateur est robuste, mais moins efficace et plus k tend vers 2, plus l'estimateur est efficace, mais moins robuste. Il suggère $p = 1, 5$ comme un bon compromis, mais ce choix n'a jamais été exploré.

3.2.1 Estimateur HB

On considère le modèle de régression linéaire ci-dessous pour lequel on suppose que, pour τ fixe, on a :

$$HB_{\tau}(\epsilon_{i,\tau} | \mathbf{x}_i) = 0 \quad (3.3)$$

où $HB_{\tau}(\epsilon_{i,\tau} | \mathbf{x}_i) = \mathbb{E}(\zeta_{\tau}^{1,5}(\epsilon_{i,\tau} | \mathbf{x}_i))$.

L'estimateur HB du vecteur des paramètres de cette régression est défini comme suit :

$$\hat{\boldsymbol{\beta}}^{HB} = \underset{\boldsymbol{\beta} \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \zeta_{\tau}^{1,5}(y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_p x_{i,p}). \quad (3.4)$$

La figure 3.1 illustre les pertes $\zeta_{\tau}^{1,5}$ pour trois valeurs de k et τ égal à 0, 5. La perte avec $k = 1, 5$ se retrouve entre les deux pertes de la régression quantile et expectile. La perte de la régression HB donne plus de poids que la RQ mais affecte moins de poids que la

RE surtout pour les grandes valeurs, c-à-d, les valeurs qui correspondent aux valeurs aberrantes. Ce qui motive le choix de l'estimation HB c'est de faire un compromis entre la robustesse et l'efficacité de RQ et de RE. Les trois courbes se croisent en $x = \pm 1$.

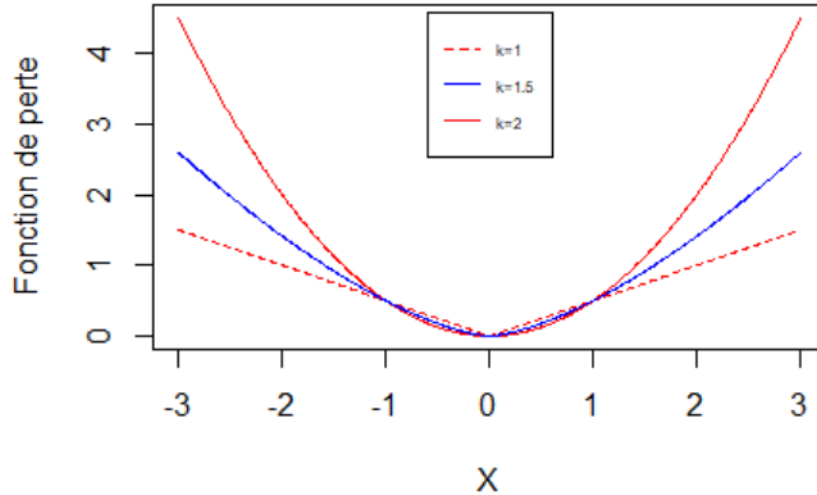


Figure 3.1 – Fonctions de perte $\zeta_{1,5}^k$ avec différentes valeurs de k .

La solution du problème d'optimisation (3.4) n'est pas analytique, car la fonction objectif dans les deux autres cas n'est pas différentiable à l'origine. Les solutions suggérées sont numériques et sont générées par différents algorithmes comme l'algorithme de sous-gradients avec contraintes.

Exemple Dans cet exemple, on considère la base de données *mtcars* étudiée par [HV81]. Cette base contient 32 observations et 11 variables. Ici, nous allons modéliser *mpg* (miles par gallon) en fonction de *wt* (weight) le poids de l'auto. Intuitivement, plus l'auto est lourde plus la consommation en essence augmente et plus le nombre de miles parcourus diminue. Pour vérifier la stabilité de notre estimateur, nous avons ajouté deux points aberrants. La figure 3.2 illustre les trois méthodes d'estimation de la régression linéaire. Nous remarquons que la régression quantile avec et sans points aberrants n'a pas changé tandis que la régression moyenne avec les MCO, la pente de la courbe a vraiment changé

et est affectée par l'ajout de ces deux points. Notre estimateur a légèrement changé ce qui indique qu'elle est plus robuste que les MCO mais moins en la comparant à la régression quantile.

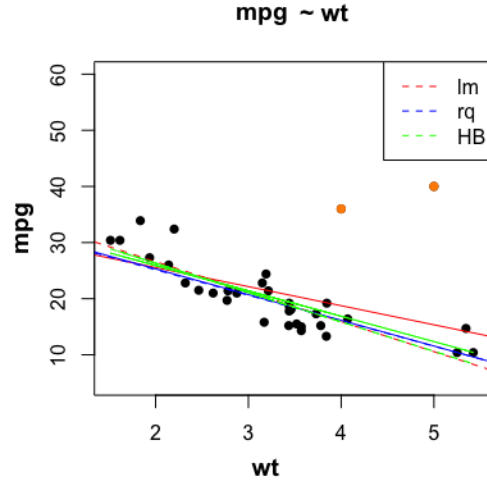


Figure 3.2 – Relation linéaire entre mpg et wt avec trois méthodes d'estimation : MCO, régression quantile et régression HB. Sans valeurs aberrantes, les droites en lignes continues et avec points aberrants les droites sont discontinues. Les deux points oranges sont des points aberrants.

3.2.2 Propriétés asymptotiques

Dans cette section, nous allons établir deux propriétés asymptotiques de l'estimateur HB : la convergence et la normalité asymptotique. Le théorème suivant montre la convergence de l'estimateur $\hat{\beta}^{HB}$

Théorème 3.1 *Quand n vers l'infini, $\hat{\beta}^{HB}$ converge en probabilité vers β_0 , où*

$$\beta_0 = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \mathbb{E} \left(\zeta_{\tau}^{1,5}(\epsilon_{\tau}) | \mathbb{X} \right).$$

Démonstration

D'abord, la loi faible des grands nombres on pour n grande

$$\frac{1}{n} \sum_{i=1}^n \zeta_{\tau}^{1,5}(y_i - \beta_0 - \beta_1 x_{i,1} - \cdots - \beta_p x_{i,p}) \xrightarrow{P} \mathbb{E}(\zeta_{\tau}^{1,5}(\epsilon_{\tau} | \mathbb{X})).$$

En plus la fonction de perte $\zeta_{\tau}^{1,5}$ est convexe. Alors, d'après le lemma 2 de Hjort et Pollard [HP93], on a la convergence en probabilité de $\hat{\beta}^{HB}$ vers β_0 . D'après ce lemma, on a aussi $\sqrt{n}(\hat{\beta}^{HB} - \beta_0) = O_p(1)$, c-à-d $\sqrt{n}(\hat{\beta}^{HB} - \beta_0)$ est borné en probabilité.

■

Pour montrer la normalité asymptotique de notre estimateur, nous avons adapté les hypothèses dans [NP87], mais la preuve est complètement différente parce que la dérivée seconde de la fonction de perte à l'origine n'existe pas.

Théorème 3.2 *Sous certaines conditions de régularité et lorsque n tend vers l'infini, nous avons :*

$$\sqrt{n}(\hat{\beta}^{HB} - \beta_0) \sim N(0, \mathbf{W}^{-1} \mathbf{V} \mathbf{W}^{-1}) \quad (3.5)$$

où $\mathbf{W} = 2\mathbb{E}\left(|\psi_{\tau}(\epsilon_{i,\tau})| \epsilon_{i,\tau}^{-\frac{1}{2}} |\mathbf{x}_i \mathbf{x}_i^t\right)$ et $\mathbf{V} = \mathbb{E}\left(|\psi_{\tau}(\epsilon_{i,\tau})| \epsilon_{i,\tau}^{-1} \mathbf{x}_i \mathbf{x}_i^t\right)$.

Démonstration

On définit la fonction multidimensionnelle suivante :

$$H(\beta) \equiv \begin{pmatrix} H_0(\beta) \\ H_1(\beta) \\ \vdots \\ H_p(\beta) \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left(\frac{\partial \zeta_{\tau}^{1,5}(\epsilon_{\tau})}{\partial \beta_0}\right) \\ \mathbb{E}\left(\frac{\partial \zeta_{\tau}^{1,5}(\epsilon_{\tau})}{\partial \beta_1}\right) \\ \vdots \\ \mathbb{E}\left(\frac{\partial \zeta_{\tau}^{1,5}(\epsilon_{\tau})}{\partial \beta_p}\right) \end{pmatrix}.$$

On applique la formule de Taylor d'ordre 1 à la fonction H au voisinage de β_0 :

$$0 = \sqrt{n}H(\beta_0) = \sqrt{n}H(\hat{\beta}) + A_{\hat{\beta}}\sqrt{n}(\hat{\beta} - \beta_0),$$

où $\tilde{\beta}$ est entre β_0 et $\hat{\beta}$ et $A_{\beta} = [A_{j,l}]$ est une matrice $(1+p) \times (1+p)$ avec

$$A_{j,l} = \mathbb{E} \left(\frac{\partial^2 \zeta_{\tau}^{1,5}(\epsilon_{\tau})}{\partial \beta_j \partial \beta_l} \right) \text{ pour } j, l = 0, 1, \dots, p,.$$

D'après la démonstration du théorème 3.1, on a $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$. En supposant la continuité de A_{β} en β , nous avons donc

$$0_{1+p} = \sqrt{n}H(\hat{\beta}) + A_{\beta_0}\sqrt{n}(\hat{\beta} - \beta_0) + o_p(1).$$

Si A_{β_0} est inversible, il vient alors :

$$\sqrt{n}(\hat{\beta} - \beta_0) = A_{\beta_0}^{-1}\sqrt{n}H(\hat{\beta}) + O_p(1).$$

Notant $\epsilon_{i,\tau} = y_i - \mathbf{x}_i^t \beta_0$ il vient :

$$\begin{aligned} H(\hat{\beta}) &= \mathbb{E} \left(-\mathbf{x}_i \frac{\partial \zeta_{\tau}^{1,5}}{\partial \beta} (Y_i - \mathbf{x}_i^t \hat{\beta}) \right) \\ &= \frac{3}{2} \mathbb{E} \left(-\mathbf{x}_i |\psi_{\tau}(\hat{\epsilon}_{i,\tau})| |\hat{\epsilon}_{i,\tau}|^{0,5} \right) \\ &= \frac{-3}{2n} \sum_{i=1}^n \mathbf{x}_i |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_{i,\tau}|^{0,5} \\ &\quad + \frac{3}{2n} \sum_{i=1}^n \mathbf{x}_i |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_{i,\tau}|^{0,5} - \frac{3}{2} \mathbb{E} \left(\mathbf{x}_i |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_{i,\tau}|^{0,5} \right) \\ &\quad + \frac{3}{2} \mathbb{E} \left(\mathbf{x}_i |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_{i,\tau}|^{0,5} \right) - \frac{3}{2} \mathbb{E} \left(\mathbf{x}_i |\psi_{\tau}(\hat{\epsilon}_{i,\tau})| |\hat{\epsilon}_{i,\tau}|^{0,5} \right) \\ &= \frac{-3}{2n} \sum_{i=1}^n \mathbf{x}_i |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_{i,\tau}|^{0,5} + I_{1,n} + I_{2,n}, \end{aligned}$$

où

$$I_{1,n} = \frac{3}{2n} \sum_{i=1}^n \mathbf{x}_i |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_{i,\tau}|^{0,5} - \frac{3}{2} \mathbb{E} \left(\mathbf{x}_i |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_{i,\tau}|^{0,5} \right)$$

et où

$$I_{2,n} = \frac{3}{2} \mathbb{E} \left(\mathbf{x}_i |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_{i,\tau}|^{0,5} \right) - \frac{3}{2} \mathbb{E} \left(\mathbf{x}_i |\psi_{\tau}(\hat{\epsilon}_{i,\tau})| |\hat{\epsilon}_{i,\tau}|^{0,5} \right)$$

Or, la loi faible des grands nombres implique que $I_{1,n} = o_p(1)$. De même :

$$|I_{2,n}| \leq \mathbb{E} \left(|\mathbf{x}_i| \left| |\epsilon_i|^{0,5} - |\hat{\epsilon}_i|^{0,5} \right| \right) = o_p(1) \quad \text{puisque} \quad \hat{\beta} - \beta_0 = o_p(1).$$

Par conséquent,

$$H(\hat{\beta}) = \frac{-3}{2n} \sum_{i=1}^n \mathbf{x}_i |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_i|^{0,5} + o_p(1).$$

Par ailleurs :

$$A_{\beta_0} = \frac{3}{4} \mathbb{E} (\mathbf{x}_i \mathbf{x}_i^t |\psi_{\tau}(\epsilon_{i,\tau})| |\epsilon_{i,\tau}|^{-0,5})$$

D'où :

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{3}{4} \mathbb{E} (\mathbf{x}_i \mathbf{x}_i^t |\psi_{\tau}(\epsilon_i)| |\epsilon_i|^{-0,5}) \right)^{-1} \frac{-3}{2\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i |\psi_{\tau}(\epsilon_i)| |\epsilon_i|^{0,5}$$

En utilisant le théorème central limite, il vient :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i |\psi_{\tau}(\epsilon_i)| |\epsilon_i|^{0,5} \xrightarrow{d} N(0, \mathbf{V})$$

où $\mathbf{V} = \mathbb{E} (\psi_{\tau}^2(\epsilon_{i,\tau}) \mathbf{x}_i \mathbf{x}_i^t \epsilon_{i,\tau})$

Donc,

$$\sqrt{n}(\hat{\beta} - \beta_0) \sim \mathcal{N}(0, \mathbf{W}^{-1} \mathbf{V} \mathbf{W}^{-1})$$

où $\mathbf{W} = 2\mathbb{E} \left(|\psi_{\tau}(\epsilon_{i,\tau})| \mathbf{x}_i \mathbf{x}_i^t |\epsilon_{i,\tau}|^{-\frac{1}{2}} \right).$

■

3.3 Estimation HB de la régression linéaire pour les données censurées

Dans cette partie nous allons introduire l'estimation HB de la régression linéaire dans le cas des données censurées. Plus exactement, on suppose que la variable d'intérêt est soumise à la censure aléatoire à droite.

En présence des données censurées, l'estimation (3.4) du modèle linéaire (3.1) n'est pas valable car la variable réponse y_i n'est pas observée complètement à cause de la censure. Dans ce cas, on observe seulement au point \mathbf{x}_i le couple de séquences i.i.d, de taille n , (T_i, δ_i) de (T, δ) , où $T = \min(Y, C)$, $\delta = \mathbb{1}(Y \leq C)$ et C est la variable de censure qui, conditionnellement à \mathbf{X} , est indépendante de Y . Nous allons procéder comme dans le chapitre 2 pour construire un nouvel estimateur de la régression linéaire, plus efficace que la RQ et plus robuste que RE.

3.3.1 Estimation de la régression HB

En présence des données censurées et en utilisant (2.3) pour une valeur fixe de $\tau \in (0, 1)$, on définit l'estimateur HB du vecteur des paramètres, $\boldsymbol{\beta}$, du modèle de la régression (3.1) comme suit :

$$\hat{\boldsymbol{\beta}}_\tau(\mathbf{x}) \stackrel{(2.3)}{=} \underset{\boldsymbol{\beta} \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \zeta_\tau^{1,5}(T_i - \xi^t \boldsymbol{\beta}) \widehat{W}_i(\mathbf{x}_i)$$

où $\zeta_\tau^{1,5}(u) = |u|^{1,5} |\psi_\tau(u)|$ avec $\psi_\tau(u) = |\tau - \mathbb{1}_{(u \leq 0)}|$ et $\widehat{W}_i(\mathbf{x}) = \frac{\delta}{1 - \widehat{G}_C(T_i^- | \mathbf{x})}$ est la fonction de poids causée par la censure, tel que $\widehat{G}_C(\cdot | \mathbf{x})$ est l'estimateur de Beran, [Ber81], de la fonction de distribution de la censure C étant donné $\mathbf{X} = \mathbf{x}$.

Comme dans le cas des données complètes, la solution n'est pas analytique et il faut donc résoudre numériquement ce problème. Dans les simulations, la fonction *optim* du logiciel *R* est utilisée pour évaluer l'estimation de $\boldsymbol{\beta}$.

Théorème 3.3 $\widehat{\boldsymbol{\beta}}^{HB}$ converge en probabilité vers $\boldsymbol{\beta}_0$ si n vers l'infini, où

$$\boldsymbol{\beta}_0 = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \mathbb{E} (\zeta_\tau^{1,5}(\mathbf{Y} - \boldsymbol{\beta} \mathbf{X}) | \mathbf{X}) .$$

Démonstration

D'abord, notons par $W_i(\mathbf{x}) = \frac{\delta}{1-G_C(T_i^-|\mathbf{x})}$. On commence par la décomposition suivante :

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \zeta_{\tau}^{1,5}(T_i - \mathbf{x}_i^t \boldsymbol{\beta}) \widehat{W}_i(\mathbf{x}) &= \mathbb{E}(\zeta_{\tau}^{1,5}(y_i - \mathbf{x}_i^t \boldsymbol{\beta}) W_i(\mathbf{x})) \\ &+ \frac{1}{n} \sum_{i=1}^n \zeta_{\tau}^{1,5}(T_i - \mathbf{x}_i^t \boldsymbol{\beta}) W_i(\mathbf{x}) - \mathbb{E}(\zeta_{\tau}^{1,5}(T_i - \mathbf{x}_i^t \boldsymbol{\beta}) W_i(\mathbf{x})) \\ &+ \frac{1}{n} \sum_{i=1}^n \zeta_{\tau}^{1,5}(T_i - \mathbf{x}_i^t \boldsymbol{\beta}) (\widehat{W}_i(\mathbf{x}) - W_i(\mathbf{x})) \end{aligned}$$

En utilisant la loi faible des grands nombres et le fait que $\sup_i |\widehat{W}_i(\mathbf{x}) - W_i(\mathbf{x})| = o_p(1)$, (sous certaines hypothèses) on a

$$\frac{1}{n} \sum_{i=1}^n \zeta_{\tau}^{1,5}(T_i - \mathbf{x}_i^t \boldsymbol{\beta}) \widehat{W}_i(\mathbf{x}) = \mathbb{E}(\zeta_{\tau}^{1,5}(y_i - \mathbf{x}_i^t \boldsymbol{\beta}) W_i(\mathbf{x})) + o_p(1).$$

En plus la fonction de perte $\zeta_{\tau}^{1,5}$ est convexe. Alors, d'après le lemme 2 de Hjort et Pollard (1992), on a la convergence en probabilité de $\hat{\boldsymbol{\beta}}^{HB}$ vers $\boldsymbol{\beta}_0$.

■

Remarque 3.4 *La normalité asymptotique est assez compliquée à cause du poids $\widehat{G}_C(\cdot|\mathbf{x})$. Établir ce résultat est laissé pour travaux future proche. Mais, par simulation, voir le chapitre suivant, nous voyons que notre estimateur suit asymptotiquement une loi normale.*

CHAPITRE 4

Simulations

Dans les précédents chapitres, nous avons introduit, dans le cas des données complètes et censurées, trois méthodes de régression : régression quantile, régression expectile et notre régression proposée, nommée la régression HB. Les propriétés asymptotiques qui montrent la performance de différentes approches d'estimation ont été présentées. Dans le présent chapitre, nous cherchons à comparer ces trois méthodes d'estimation pour des données de taille finie. Pour cela, nous générons des échantillons à partir des modèles linéaires dans différents scénarios : données complètes et censurées, erreurs homoscedastiques et hétéroscedastiques et avec ou sans des données aberrantes. À l'aide des résultats de simulations, et afin de prouver l'efficacité et la robustesse de notre nouvel estimateur HB, nous évaluons la qualité des estimateurs étudiés dans ce travail en termes de biais et d'écart type. Nous allons aussi vérifier la normalité des estimateurs grâce à ces simulations.

4.1 Les Modèles et leurs paramètres

Dans cette section, nous considérons différents modèles pour effectuer les simulations dans le cas des données complètes et censurées. Nous présentons dans ces deux cas, deux sortes de modèles linéaires, des modèles linéaires sans valeurs aberrantes pour montrer l'efficacité des estimateurs et des modèles linéaires ayant des données aberrantes pour lesquels nous évaluons la robustesse des estimateurs.

Données complètes

Dans cette partie, nous présentons les différents modèles linéaires qui ont été utilisés pour les simulations dans le cas des données complètes. Ces modèles sont divisés en deux parties. La première partie est dédiée aux modèles linéaires sous l'hypothèse d'homoscédasticité et la deuxième aux modèles linéaires sous l'hypothèse d'hétéroscédasticité. Pour tous les modèles proposés, les erreurs ϵ_i sont générées aléatoirement à partir d'une loi normale centrée réduite ($\mu = 0, \sigma = 1$). Pour vérifier la sensibilité des estimateurs, on considère les mêmes modèles, mais avec des points aberrants fixes, et non pas générés aléatoirement.

Remarques :

1. Sous l'hypothèse d'homoscédasticité du modèle $y = \beta_0 + x\beta_1 + \epsilon$, le quantile de la variable d'intérêt est : $q_\tau(y|x) = \beta_0^\tau + x\beta_1^\tau$ où $\beta_0^\tau = \beta_0 + q_\tau(\epsilon_\tau)$. De sorte que, seule l'ordonnée à l'origine est fonction de τ et les fonctions quantiles sont des lignes parallèles.
2. Sous l'hypothèse d'hétéroscédasticité, par exemple $y = \beta_0 + x\beta_1 + x\epsilon$, le quantile de la variable réponse y est donné par $q_\tau(y|x) = \beta_0^\tau + x\beta_1^\tau$ où cette fois-ci la pente

$\beta_1^\tau = \beta_1 + q_\tau(\epsilon_\tau)$ est fonction du paramètre τ .

Ces remarques s'appliquent également au modèle de la RQ et de la HB. Donc, il faut réajuster l'estimation du paramètre β dans le cas d'hétéroscédasticité. A l'aide de la simulation en logiciel R, on a pu récupérer les valeurs des quantiles, expectiles et la fonction $\zeta_\tau^{1.5}(\epsilon)$ pour $\epsilon \sim N(0, 1)$ avec différentes valeurs de $\tau = 0.25, 0.5$ et 0.75 . Notons que les valeurs sont symétriques.

— Pour le quantile on a : $(q_{0.25}(\epsilon), q_{0.5}(\epsilon), q_{0.75}(\epsilon)) = (-0.6744898, 0, 0.6744898)$.

— Pour l'expectile on a : $(\mu_{0.25}(\epsilon), \mu_{0.5}(\epsilon), \mu_{0.75}(\epsilon)) = (-0.4363266, 0, 0.4363266)$.

— Pour la régression HB on a : $(\zeta_{0.25}^{1.5}(\epsilon), \zeta_{0.5}^{1.5}(\epsilon), \zeta_{0.75}^{1.5}(\epsilon)) = (-0.5207031, 0, 0.5207031)$.

Modèle 1 : $Y_i = 1 + 2X_i + \epsilon_i$, $i = 1, \dots, n$.

Le modèle 1 est un modèle linéaire simple sous l'hypothèse d'homoscédasticité c-à-d la variance de l'erreur des variables est constante. Dans notre modèle, la variance est égale à 1 et où la variable explicative X suit une loi normale centrée réduite ; $X_i \sim \mathcal{N}(0, 1)$, $i = 1, \dots, n$.

Modèle 2 (modèle 1 avec un point aberrant) : $Y_i = 1 + 2X_i + \epsilon_i$, $i = 1, \dots, n$.

Le modèle 2 est le même modèle linéaire 1 mais avec un point aberrant, $(x, y) = (5, -20)$.

Modèle 3 : $Y_i = 1 + 2X_i + X_i\epsilon_i$, $i = 1, \dots, n$.

Le modèle 3 est un modèle linéaire sous l'hypothèse d'hétéroscédasticité c-à-d les variances des résidus des variables examinées sont différentes, dans notre modèle, la variance pour l'observation Y_i , conditionnelle à $\mathbf{X}_i = x_i$, est égale à $x_i^2\sigma^2$. Les valeurs de la variable explicative X sont générées de la loi gamma ; $x_i \sim \text{Gamma}(2, 3)$, $i = 1, \dots, n$.

Modèle 4 (modèle 3 avec un point aberrant) : $Y_i = 1 + 2X_i + X_i\epsilon_i$, $i = 1, \dots, n$. Le modèle 4 est le même modèle linéaire 3 mais avec un point aberrant, $(x, y) = (5, -20)$.

Données censurées

Dans cette partie, nous présentons dans le cas des données censurées, le modèle linéaire effectué lors de la simulation.

Modèle 5 : On considère le modèle linéaire suivant : Dans un premier temps, il faut savoir que la variable explicative X_i suit une loi uniforme, $X_i \sim \text{Uniforme}(0, 1)$, alors que la variable endogène Y_i est de forme linéaire,

$$Y_i = 2 + 4X_i + \epsilon_i, \quad i = 1, \dots, n.,$$

où

$$\epsilon \sim \mathcal{N}(0, 0.1).$$

Par ailleurs, la variable des temps de censure C_i suit une loi uniforme :

$$C_i \sim \text{Uniforme}(0, \lambda)$$

où λ est définie pour lequel le niveau de censure $p = \mathbb{P}(C \leq Y)$. Ainsi, la variable "réponse" observée T_i est égale à

$$T_i = \min(Y_i, C_i)$$

et la variable indicatrice des temps de censure δ_i est

$$\delta_i = \mathbb{1}(T_i \leq C_i).$$

La forme de l'échantillon final utilisée pour comparer les trois méthodes d'estimation dans le cas des données censurées est alors celle qui suit :

$$(T_i, \delta_i, X_i), \quad i = 1, \dots, n.$$

Afin de comparer les trois méthodes, dans les deux cas de données complètes et censurées, décrites aux précédents chapitres, plusieurs échantillons ont été générés 1000 fois à partir des modèles présentés en haut. Plusieurs tailles d'échantillon ($n = 50, n = 100$ et $n = 200$) et différentes valeurs de $\tau = (0.25, 0.5, 0.75)$ ont été considérés. En plus, pour les données censurées, on considère de différents niveaux de censure ($p = 0, 25, 0, 5$). Les trois méthodes de régression et les données ont toutes été programmées et simulées à partir du logiciel *R*.

4.2 Les résultats

Dans cette section, les résultats obtenus pour les trois méthodes de régression avec variable réponse complète et censurée à droite vues aux précédents chapitres seront présentés. Les résultats seront donnés sous forme de table pour les différentes valeurs des paramètres discutés à la section 4.1. Par la suite, une discussion sera faite à partir ces résultats.

D'abord et avant de présenter les résultats de simulations qui comparent les performances des trois estimateurs, nous allons vérifier la normalité de l'estimateur HB.

4.2.1 Normalité de l'estimateur HB

Nous avons démontré la normalité asymptotique dans le cas de données complètes. Pour se faire, nous avons utilisé les 1000 valeurs de l'estimateur de β par la méthode HB pour construire l'estimateur non-paramétrique à noyau de $\hat{\beta}$. Nous avons considéré la densité gaussienne comme noyau et le paramètre de lissage est sélectionné par la méthode de validation-croisée. Dans le logiciel *R*, l'estimation non-paramétrique est donnée par la

fonction *density*. Ainsi, les densités de $\hat{\beta}$ pour différentes valeurs de $\tau = (0.25, 0.5, 0.75)$ sont illustrées aux figures 4.1-4.4 pour le modèle 1, 3 et 5 avec le degré de censure $p = 0.25$ et $p = 0.5$ respectivement.

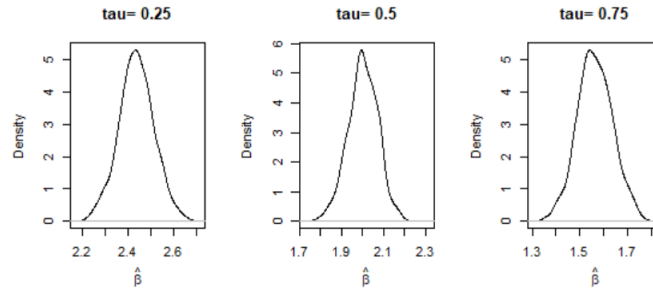


Figure 4.1 – Densité de l’estimateur HB du modèle 1 pour de différentes valeurs de τ où $n = 200$.

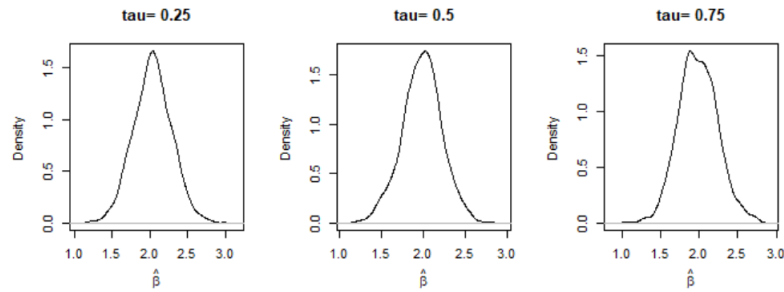


Figure 4.2 – Densité de HB du modèle 3 pour différentes valeurs de τ où $n = 200$.

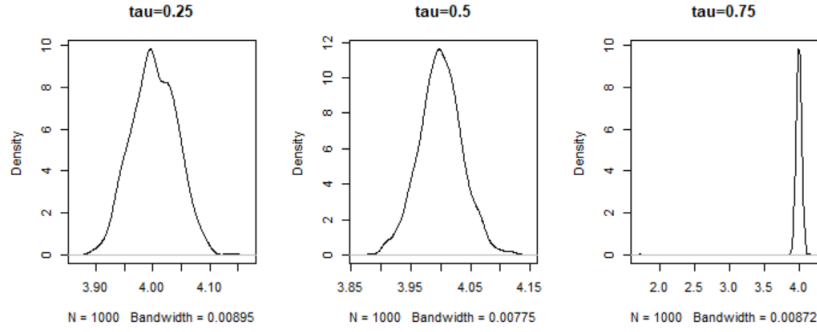


Figure 4.3 – Densité de l’estimateur HB du modèle 5 pour de différentes valeurs de τ où $n = 200$ dans le cas de censure à droite où $p = 0, 25$.

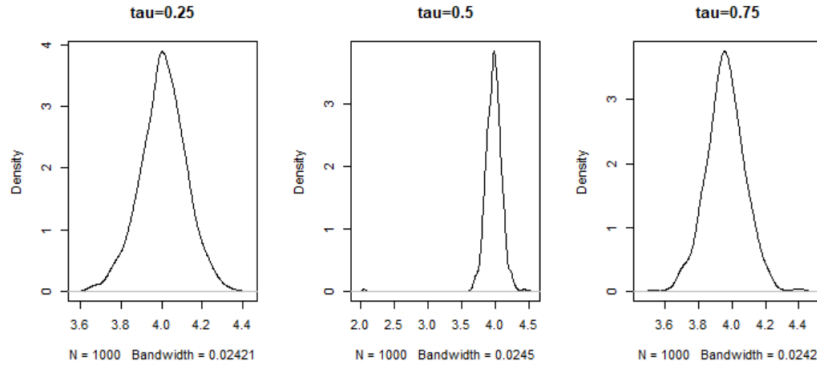


Figure 4.4 – Densité de l’estimateur HB du modèle 5 pour de différentes valeurs de τ où $n = 200$ dans le cas de censure à droite où $p = 0, 5$.

À partir des figures 4.1, 4.2, 4.3 et 4.4, on peut voir que, pour les modèles 1, 3 et 5, les distributions de l’estimateur $\hat{\beta}$ sont symétriques, autour de 2 pour le modèle 1 et 3 et autour de 4 pour le modèle 5 où $p = 0.25$ et 0.5 . La distribution de $\hat{\beta}$ semble suivre une loi normale. Ce qui confirme les résultats théoriques démontrés pour le cas des données complètes.

4.2.2 Comparaison des estimateurs

Après la présentation des modèles et leurs paramètres, les résultats de chacune des trois méthodes peuvent maintenant être examinés et comparés. Ainsi, les tableaux qui suivent montrent le biais et l'écart type obtenus lors de l'estimation des modèles développés à la section 4.1.

Ces tableaux sont donc illustrés ci-dessous et, pour chacun d'eux, une courte de table comparant le biais et l'écart-type suit. Une discussion à propos de l'ensemble de ces tableaux sera par la suite présentée à la section 4.2.3.

Tableau 4.1 – Table présentant le carré du biais (10^{-3}) et les écarts types (10^{-2}) obtenus à partir du modèle 1 de chacune des trois méthodes pour $n = 50, 100$ et 200 et $\tau = 0, 25, 0, 5$ et $0, 75$.

Modèle 1							
τ	n	Quantile		HB		Expectile	
		Biais ²	SD	Biais ²	SD	Biais ²	SD
$\tau = 0.25$	n=50	0.049345	19.2854	0.029593	15.5843	0.011588	14.9564
	n=100	0.076207	13.69531	0.054434	11.4529	0.041412	10.8690
	n=200	0.006202	9.7305	0.004918	7.9416	0.004344	7.6487
$\tau = 0.5$	n=50	0.041176	18.1469	0.019898	15.1852	0.018355	14.2649
	n=100	0.072949	12.48612	0.066389	10.5705	0.026409	10.3090
	n=200	0.009358	9.086201	0.002557	7.428582	0.001514	7.1341
$\tau = 0.75$	n=50	0.079463	19.1329	0.060297	15.4281	0.063875	14.8267
	n=100	0.025469	14.10643	0.013830	11.4626	0.008046	10.9511
	n=200	0.005535	9.683096	0.001281	7.7805	0.000140	7.32961

Dans la table 4.1 ci-dessus, qui présente le carré du biais et le SD obtenus par les trois différentes méthodes d'estimation du paramètre $\beta = 2$ du modèle 1. D'abord, dans ce cas idéal c-à-d un modèle linéaire homoscédastique, le biais et le SD sont très petits même pour une taille d'échantillon $n = 50$. On voit que la méthode de HB pour les différentes valeurs de τ produit toujours un biais et SD qui est entre le biais et le SD de la méthode de quantile et d'expectile. On peut constater aussi que plus la taille de l'échantillon augmente plus les valeurs du biais et SD diminuent. Ce qui montre que le biais et la variance convergent vers zéro quand la taille de l'échantillon tend vers l'infini.

Tableau 4.2 – Table présentant le carré du biais (10^{-3}) et les écarts types (10^{-2}) obtenus à partir du modèle 2 de chacune des trois méthodes pour $n = 50, 100$ et 200 et $\tau = 0, 25, 0, 5$ et $0, 75$.

Modèle 2							
τ	n	Quantile		HB		Expectile	
		Biais ²	SD	Biais ²	SD	Biais ²	SD
$\tau = 0.25$	n=50	74.35463	20.86829	2406.46344	36.83562	8656.18211	32.69511
	n=100	17.35704	14.50337	443.79994	14.74106	3951.88553	20.73372
	n=200	3.709445	9.881362	93.212884	8.735887	1402.847878	11.719245
$\tau = 0.5$	n=50	17.6159	18.37630	495.9322	20.53701	4469.4005	29.74583
	n=100	4.224613	12.79367	111.034977	11.73458	1592.751008	16.76648
	n=200	0.726324	8.970634	25.2663288	7.472628	478.0756648	8.941618
$\tau = 0.75$	n=50	6.396174	19.33643	158.877495	17.82964	2054.568964	26.82289
	n=100	1.126611	13.90021	38.694129	11.62605	645.462189	13.94202
	n=200	0.451734	9.670543	10.116432	7.785984	187.001666	8.060441

La table 4.2 présente les valeurs du carré du biais et de SD engendrés par les différentes méthodes d'estimation pour le modèle 2 pour $n = 50, 100$ et 200 et $\tau = (0.25, 0.5, 0.75)$. Comparées aux résultats dans la table 4.1, toutes les valeurs pour le modèle 1 augmentent et même explosent pour l'expectile. La régression quantile est la meilleure au niveau du biais suivi par la régression HB. La régression expectile est clairement très sensible au point aberrant ajouté à l'échantillon. Par contre, au niveau de l'écart type, la méthode HB a en général fournit les plus petites valeurs, et ce pour $\tau = 0.5$ et $\tau = 0.75$. On constate aussi, quand $n = 200$, le biais et l'écart type sont plus faibles en comparant les deux autres valeurs de n . On voit aussi, que le point aberrant a non seulement affecté le biais, mais la stabilité de la régression expectile.

Tableau 4.3 – Table présentant le carré du biais (10^{-3}) et les écarts types (10^{-2}) obtenus à partir du modèle 3 de chacune des trois méthodes pour $n = 50, 100$ et 200 et $\tau = 0, 25, 0, 5$ et $0, 75$.

Modèle 3							
τ	n	Quantile		HB		Expectile	
		Biais ²	SD	Biais ²	SD	Biais ²	SD
$\tau = 0.25$	n=50	2.811675	38.11430	3.431535	39.43492	6.633633	46.30177
	n=100	0.4983948	26.08651	0.7057442	27.54918	1.6734542	34.04644
	n=200	0.3207828	18.39797	0.4585546	19.98662	0.8194694	25.58516
$\tau = 0.5$	n=50	0.01599888	35.18997	0.02354715	37.44139	0.02727494	45.14804
	n=100	0.00209367	23.32700	0.0066359	25.75094	0.01475997	32.76229
	n=200	0.000728	16.58288	0.001026	0.00181375	0.01022267	23.79686
$\tau = 0.75$	n=50	1.518649	38.11707	2.056806	39.59204	4.707310	46.88412
	n=100	0.7076478	26.75653	0.9247655	28.45715	2.0804715	35.00842
	n=200	0.04390425	17.84414	0.13034427	19.81498	0.45466164	25.34253

La table 4.3 présente les résultats pour le modèle linéaire sous l'hypothèse d'hétéroscédasticité. Le biais et le SD du modèle 3 donnés par la méthode HB sont entre les valeurs de l'estimation par la régression quantile et expectile. Mais cette différence entre la régression quantile et HB est très petite au niveau du biais et les deux dominent l'estimation. Au niveau de la variance, on remarque que, comparées au modèle 1, les valeurs de SD ont augmenté. Aussi, les valeurs SD de notre estimateur sont légèrement supérieures aux valeurs de SD de l'estimateur de la régression quantile. En outre, l'estimation HB domine la méthode d'estimation par la régression expectile au niveau la variance. Enfin, on constate que le carré du biais et l'écart-type décroissent avec la taille de l'échantillon. Ce qui est rassurant aux vues statistiques puisque plus le nombre d'observations est grand, plus l'écart-type et le biais sont petits.

Tableau 4.4 – Table présentant le carré du biais et les écarts types (10^{-2}) obtenus à partir du modèle 4 de chacune des trois méthodes pour $n = 50, 100$ et 200 et $\tau = 0, 25, 0, 5$ et $0, 75$.

Modèle 4							
τ	n	Quantile		HB		Expectile	
		Biais ²	SD	Biais ²	SD	Biais ²	SD
$\tau = 0.25$	n=50	0.942158	55.75415	19.78809	50.28991	25.56299	30.71447
	n=100	0.13484	27.45288	5.993389	43.15629	17.15390	31.86106
	n=200	0.027323	17.79534	1.159662	22.12379	9.066792	27.27057
$\tau = 0.5$	n=50	0.198851	37.79681	7.379896	66.53168	20.532258	53.26196
	n=100	0.044057	24.95129	1.611617	34.66319	10.947217	43.38201
	n=200	0.01088269	16.58479	0.385294	20.19370	4.700560	29.43242
$\tau = 0.75$	n=50	0.089375	41.29893	2.759234	62.14425	14.51397	71.85731
	n=100	0.020303	27.46915	0.645492	35.33242	6.454428	50.37421
	n=200	0.0055597	18.16783	0.1562759	21.98906	2.4325928	30.87067

La table 4.4 présente les résultats pour le modèle linéaire sous l'hypothèse d'hétéroscédasticité avec un point aberrant. Comme les autres tables vues jusqu'à présent dans cette section, cette dernière permet de constater que l'estimation de HB produit des biais et des SD entre celles des autres méthodes (quantile et expectile). En revanche, le biais et l'écart type observés pour ce modèle sont beaucoup plus grands, à cause du point aberrant et l'hypothèse d'hétéroscédasticité, en comparant avec tous les autres modèles présentés auparavant. Cette augmentation est constatée surtout pour la méthode d'expectile suivie par notre approche. Par contre, la méthode quantile reste toujours plus robuste avec un biais plus petit.

Dans les prochaines tables, on s'intéresse à l'analyse dans le cas des données censurées, où le taux de censure $p = 0.25$ et 0.5 pour de différentes valeurs de τ et n .

Tableau 4.5 – Table présentant le carré du biais ($\div 10^5$) et les écarts types (10^{-2}) obtenus à partir du modèle 5 de chacune des trois méthodes pour $n = 50, 100$ et 200 , $\tau = 0, 25, 0, 5$ et $0, 75$ et $p = 0, 25$.

Modèle 5, $p = 0.25$							
τ	n	Quantile		HB		Expectile	
		Biais ²	SD	Biais ²	SD	Biais ²	SD
$\tau = 0.25$	n=50	0.10411343	8.111229	0.06881993	6.649709	0.05667720	6.345875
	n=100	0.09467532	7.737777	0.0590849	4.041669	0.04870192	4.214422
	n=200	0.0386792516	3.932200	0.0127615277	3.146248	0.003113954	3.001006
$\tau = 0.5$	n=50	0.12691371	9.269258	0.11591491	6.024404	0.09523147	5.839246
	n=100	0.07209421	7.737777	0.057972604	4.214422	0.04057323	4.041669
	n=200	0.05022458	6.589432	0.04621740	2.979122	0.0396238	2.849168
$\tau = 0.75$	n=50	3.782469	7.991688	2.189422	6.503118	1.901776	6.107826
	n=100	1.69985	5.691577	0.999501	4.558391	0.6712978	4.289190
	n=200	0.646553	3.917381	0.2897956	3.124365	0.1555250	2.957588

Tableau 4.6 – Table présentant le carré du biais ($\div 10^5$) et les écarts types (10^{-2}) obtenus à partir du modèle 5 de chacune des trois méthodes pour $n = 50, 100$ et 200 , $\tau = 0, 25, 0, 5$ et $0, 75$ et $p = 0, 5$.

Modèle 5, $p = 0.5$							
τ	n	Quantile		HB		Expectile	
		Biais ²	SD	Biais ²	SD	Biais ²	SD
$\tau = 0.25$	n=50	3.280819	10.580588	1.565982	8.707096	1.442885	8.241278
	n=100	0.04879196	10.134070	0.02436582	5.929846	0.01711652	5.559552
	n=200	0.006178609	8.300947	0.003452376	4.015209	0.00211149	3.797385
$\tau = 0.5$	n=50	0.06382147	9.696529	0.01983708	8.021023	0.01241193	7.641070
	n=100	0.006684848	6.870348	0.003178940	5.706684	0.001859965	5.516693
	n=200	0.00450354	3.603872	0.00308303	3.016479	0.001529899	2.885008
$\tau = 0.75$	n=50	5.989486	10.518427	2.588772	8.426494	1.992264	7.975530
	n=100	1.5757801	7.384044	0.6813046	5.982541	0.4518323	5.695802
	n=200	0.0625401	5.391209	0.0577093	4.395630	0.0163806	4.122587

Les tables 4.5 et 4.6 présentent les résultats du biais et SD pour les trois estimateurs dans le cas des données censurées. La table 4.5 utilise le modèle 5 avec le degré de censure égal à $p = 0.25$ alors que la table 4.6 est avec $p = 0.5$. Ceci a pour but de voir l'effet du degré de la censure sur la qualité de l'estimation par les trois méthodes. D'abord, on remarque que le biais et la variance diminuent avec la taille de l'échantillon et ceci pour toutes les valeurs de τ et p . Aussi, ces deux tables donnent, en général, les mêmes conclusions tirées des tables précédentes. En effet, la méthode de HB produit un biais et un écart type entre celles des autres méthodes. En fait, il est plus stable que la régression quantile, mais moins que la régression expectile. On remarque que la régression expectile domine au niveau du biais suivi par l'estimation HB. Cependant, la différence au niveau de biais n'est pas si importante. Finalement, comme attendu, plus le degré de censure augmente plus la qualité des trois estimateurs diminue.

4.2.3 Discussion

Les simulations basées sur ces modèles confirment l'efficacité de la régression expectile en absence des valeurs aberrantes, mais sa sensibilité à un seul point aberrant. Aussi, l'estimation de la régression quantile donne des estimateurs plus stables et moins affectés, ou pas du tout, par la présence de valeurs extrêmes. Cependant, la stabilité de cette méthode est dominée par celle de la régression expectile. Notre méthode d'estimation des paramètres de la régression linéaire fournit des estimations plus efficaces que la régression quantile et plus stable que la régression expectile. Pour trouver les estimateurs dans les simulations, nous avons utilisé la fonction *optim* dans le logiciel R. Cette fonction n'est pas optimale pour trouver le minimum de la fonction de perte non-différentiable, ce qui est le cas de notre méthode et pour la régression quantile et peut-être un avantage pour la régression expectile. Développer des algorithmes d'optimisation pour notre méthode, en adaptant par exemple les algorithmes de la régression quantile, serait intéressant surtout avec les résultats prometteurs de notre méthode.

CONCLUSION

Dans ce mémoire, nous avons introduit, dans le cas des données complètes et censurées, une nouvelle classe d'estimateurs plus efficace que la RQ et plus robuste que la RE ; appelée régression HB. Cette nouvelle méthode estime l'influence des régresseurs sur tous les percentiles de la distribution conditionnelle de la variable réponse. La performance au niveau de l'efficacité et la robustesse, de l'estimation de cette nouvelle méthode de régression est entre celle de la régression quantile et de la régression expectile.

Afin de permettre une meilleure compréhension de cette méthode, dans un premier temps, nous avons étudié dans le cas des données complètes, les méthodes de régression pour différentes fonctions de perte, symétriques et asymétriques. Par la suite, nous avons présenté la RQ et la RE, dans le cas des données censurées. Dans le chapitre trois, nous avons introduit la régression linéaire HB, ses estimateurs et ses propriétés asymptotiques dans le cas des données complètes et censurées. Finalement, dans le chapitre de simulations, nous avons comparé les RQ, RE et HB à partir de leurs biais et écart type. Les résultats sont très satisfaisants et laissent entendre que notre estimateur proposé est bel est bien un estimateur efficace et robuste. En effet, les résultats des biais et des écarts types de la régression HB sont toujours entre ceux de la RQ et RE. Ceci pour les différents modèles proposés dans ce chapitre.

Dans des travaux futurs, nous pouvons étudier plusieurs pistes. Premièrement, nous pour-

rions établir les résultats théoriques de notre méthode pour les données censurées. Deuxièmement, nous pourrions développer des algorithmes d'optimisation pour trouver les estimateurs des paramètres de la nouvelle méthode de modélisation de la régression linéaire. Troisièmement, il sera intéressant d'étudier cette méthode dans le cas où le nombre de variables explicatives est supérieur au nombre d'observations. Une sélection des variables en adaptant la régression LASSO dans ce cas est fortement suggérée. Quatrièmement, dans ce travail et dans le cadre de la censure, nous avons supposé que seulement la variable réponse est censurée. Il arrive parfois que certaines variables explicatives soient également censurées et notre méthode doit être adaptée pour modéliser ce type des données. Cinquièmement, il existe d'autres formes de données incomplètes, et parmi ces cas on trouve les données manquantes et les données tronquées. Aussi, dans le présent travail, nous avons supposé que les observations sont i.i.d et pas des données dépendantes du temps comme les séries chronologiques. Adapter notre méthode et construire de nouvelles méthodes d'estimation pour la régression linéaire avec ces données serait d'un grand intérêt. Finalement, nous avons supposé, dans ce travail, une relation linéaire entre la variable réponse et les variables explicatives. Relaxer cette hypothèse et étudier la nouvelle approche d'estimation avec un modèle non-paramétrique sera une des idées à développer plus tard, et ceci pour les données mentionnées précédemment.

Bibliographie

- [And08] R. Andersen. *Modern Methods For Robust Regression*. SAGE Publications, 2008.
- [AR95] B. Abdous and B. Rémillard. Relating quantiles and expectiles under weighted-symmetry. *Annals of the Institute of Statistical Mathematics*, pages 371–384, 1995.
- [Azz85] A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(4) :171–178, 1985.
- [Ber81] R. Beran. Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley, 1981.
- [BJ79] J. Buckley and I. James. Linear regression with censored data. *Biometrika*, pages 429–436, 1979.
- [CT13] C. Chenlei and X. Tong. A quantile regression estimator for censored data. *Bernoulli*, 19(1) :344–361, 02 2013.
- [CW05] C. Chen and Y. Wei. Expectiles and m-quantiles are quantiles. *Sankhya : The Indian Journal of Statistics*, pages 399–417, 2005.
- [DH83] D.L. Donoho and P.J. Huber. *The notion of breakdown point*. In A Festschrift for Erich Lehmann (P. Bickel and K. Doksum and J.L. Hodges, Jr.), Wadsworth, Belmont, 1983.

- [DS98] N.R. Draper and H. Smith. *Applied Regression Analysis*. Wiley Interscience Publication, Third edition, 1998.
- [Efr91] B. Efron. Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, pages 93–125, 1991.
- [FG96] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman and Hall., New York, NY., 1996.
- [FR13] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Basel, Switzerland : Birkhäuser, 2013.
- [FW06] B. Fitzenberger and R. A. Wilke. Using quantile regression for duration analysis. *Allgemeines Statistisches Archiv*, 90(1) :105–120, 2006.
- [HBK16] H. Holzmann and B. Bernhard Klar. Expectile asymptotics. *Electronic Journal of Statistics*, pages 2355–2371, 2016.
- [HM85] W. Härdle and J.S. Marron. Optimal bandwidth selection in nonparametric regression function. *Ann. Statist.*, 13 :1465–1481, 1985.
- [HP93] N.L. Hjort and D. Pollard. *Asymptotics for Minimisers of Convex Processes*. Statistical research report. Universitetet i Oslo. Matematisk Institutt, 1993.
- [HT15] T. Hastie and R. Tibshirani. *Statistical learning with sparsity : the lasso and generalizations*. Monographs on statistics and applied probability, no. 143. CRC Press, special indian ed. edition, 2015.
- [Hub64] P.J. Huber. *Robust Estimation of a Location Parameter*, volume 35. The Annals of Mathematical Statistics, 1964.
- [Hub81] P.J. Huber. *Robust Statistics*. Wiley, 1981.
- [HV81] H.V. Henderson and P.F. Velleman. Building multiple regression models interactively. *Biometrics*, 37 :391–411, 1981.

- [Jon94] M.C. Jones. Expectiles and m-quantiles are quantiles. *Statistics and Probability Letters*, pages 149–153, 1994.
- [JSP13] J. Jureckova, P.K. Sen, and J. Picek. *Methodology in Robust and Nonparametric Statistics*. CRC Press, 2013.
- [KB78] R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46(1) :33–50, 1978.
- [KM58] E. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53 :457–481, 1958.
- [KM03] J.P. Klein and M.L. Moeschberger. *Survival Analysis Techniques for Censored and Truncated Data*. Second edition, 2003.
- [Koe05] R. Koenker. *Quantile Regression*. Cambridge University Press, 2005.
- [Mil76] R. Miller. Least squares regression with censored data. *Biometrika*, 63 :449–464, 1976.
- [NP87] W.K. Newey and J.K Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55(4) :819–847, 1987.
- [PK97] S. Portnoy and R. Koenker. The gaussian hare and the laplacian tortoise : computability of squared-error versus absolute-error estimators. *Statist. Sci.*, 12(4) :279–300, 11 1997.
- [RL87] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. Hoboken :Wiley, 1987.
- [RY84] P. Rousseeuw and V.J. Yohai. Robust Regression by Means of S-Estimators I. *Robust and Nonlinear Time Series Analysis*, pages 256–272, 1984.
- [SPHL13] Y. Susanti, H. Pratiwi, S. Sulistijowati H., and T. Liana. M Estimation, S Estimation, and MM E estimation In Robust Regression. *International Journal of Pure and Applied Mathematics*, pages 349–360, 2013.

- [Sto77] C. J. Stone. Consistent nonparametric regression. *Ann. Statist.*, 5(4) :595–645, 1977. With discussion and a reply by the author.
- [YJ98] K. Yu and M. C. Jones. Local linear quantile regression. *Journal of the American Statistical Association*, pages 228–237, 1998.
- [Yoh87] V. J. Yohai. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(20) :642–656, 1987.
- [ZWZ09] V. Zinde-Walsh and D. Zhu. Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, 148 :86–99, 02 2009.
- [ZYW95] S. H. Jung Z. Ying and L. J. Wei. Survival analysis with median regression models. *Journal of the American Statistical Association*, 90 :178–184, 1995.